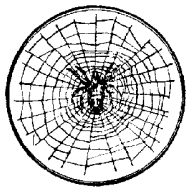
ŽILINSKÁ UNIVERZITA
FAKULTA RIADENIA A INFORMATIKY



Doc. RNDr. Stanislav Palúch, CSc.

TEÓRIA INFORMÁCIE

VYDALA ŽILINSKÁ UNIVERZITA V ŽILINE, 2008
Posledná úprava: 29. januára 2008 (1092. minúta dňa)

Vedecký redaktor:

Recenzenti: Prof. RNDr. Ján Černý, Dr.Sc., DrHc.
Prof. RNDr. Beloslav Riečan, Dr.Sc.

Vydala Žilinská univerzita v Žiline – EDIS vydavateľstvo ŽU

ISBN-80-8070-XXX-X

– Kde je moudrost?
– Ztracena v znalostech.
– Kde jsou znalosti?
– Ztraceny v informacích.
– Kde jsou informace?
– Ztraceny v datech.

Anonymus.

Úvod

So vstupom do tretieho tisícročia vstupuje ľudstvo do veku, ktorý sa bude právom charakterizovať prívlastkom *informačný*. Sme a stále viac budeme zahrňovaní množstvom informácií najrôznejšieho druhu. Tlač, televízia, rozhlas so svojimi pozemskými i satelitnými verziami a v poslednej dobe hlavne internet sú zdrojmi stále väčšieho množstva informácie. Množstvo informácií vzniká, prenáša sa a spracováva sa v súvislosti s činnosťou štátnych a regionálnych úradov, výrobných podnikov, bánk, poisťovní rôznych fondov, škôl, zdravotníckych zariadení, polície, bezpečnostných služieb, i samotných občanov. Najčastejším operáciami s informáciou je jej prenos a ukladanie, spracovanie a využívanie. V poslednej dobe rastie tiež význam jej ochrany pred odcudzením, zneužitím či neautorizovanou zmenou.

Technológia prenosu, ukladania a spracovávania informácie podstatne ovplyvňuje rozvoj ľudskej civilizácie. Literatúra uvádza ako prvú informačnú revolúciu vynájdenie písma. Dovtedy ústne odovzdávaná informácia sa uložením dala prenášať v priestore i čase. To spôsobilo, že civilizácie ovládajúce písmo začali predbiehať dovtedy rovnako rozvinuté spoločenstvá – dodnes sa nájdu niektoré kmene v zabudnutých častiach sveta stále žijúce v dobe kamennej.

Druhú informačnú revolúciu spôsobilo vynájdenie kníhtlače. Možnosť šírenia informácie medzi široké vrstvy ľudí spôsobilo enormný nárast vzdelanosti a dalo základy pre priemyselnú revolúciu a vznik modernej industriálnej spoločnosti a viedlo k súčasnej vedeckotechnickej revolúcii.

Tretia informačná revolúcia sa spája s rozvojom výpočtovej a komunikačnej techniky a s jej schopnosťou ukladať, prenášať a spracovávať enormné množstvo informácie. Oslobodenie informácie od jej materiálneho nosiča pri prenose

a obrovská kapacita pamäťových médii spolu s počítačovým spracovaním sa považuje za nástroj rozvoja s nedoziernymi dôsledkami.

Na druhej strane je však súčasná rozvinutá spoločnosť oveľa zložitejšie organizovaná. Globalizácia sveta je jedným z charakteristických javov súčasnosti. Ekonomiky jednotlivých krajín už nie sú izolované – charakteristické sú nadnárodné spoločnosti. Dnes pravdepodobne neexistuje zložitejší výrobok, ktorý by bol vyrobený len v jednej krajine. Podstatné problémy krajín prerastajú ich hranice a stávajú sa celosvetovými. Takými sú ochrana životného prostredia, globálne otepľovanie, jadrová energetika, nezamestnanosť, medzinárodná kriminalita atď. Riešenie takýchto problémov vyžaduje súčinnosť vlád, manažmentov veľkých podnikov i správ väčších a menších regionálnych celkov, miest, obcí i samotných občanov, čo je nemožné bez prenosu informácií medzi jednotlivými subjektami.

Jednou z úlohu každej modernej spoločnosti je teda budovať dostatočne výkonnú sieť na prenos informácií a optimálne ju využívať. Výstavba spojovacích sietí je však veľmi nákladná a preto často stojíme pred otázkou, či je už dané spojenie využité na maximum kapacity, alebo sa dá použitím nejakej optimalizačnej metódy preniesť cez toto spojenie viac informácie.

Dať fundovanú odpoveď na túto otázku nebolo a dodnes nie je ľahké. Použitie optimalizačnej metódy vyžaduje vytvoriť matematický model zdroja informácie, prenosovej cesty, dejov a procesov, ktoré sa odohrávajú pri prenose informácie. Tieto problémy sa začali veľmi nástojčivo ohlasovať po II. svetovej vojne. Nebolo ich možné zaradiť do žiadnej dovtedy etablovanej matematickej disciplíny. Musel teda vzniknúť nový vedný odbor – teória informácie. Tá sa stala súčasťou vtedy vznikajúcej vedy o riadení – matematickej kybernetiky, ktorá postupným vývojom prerástla do ešte mladšej vedeckej disciplíny – informatiky.

Teória informácie delí prenos informácie na nasledujúce fázy:

- vysielanie správ zo zdroja
- kódovanie správ v kóderi
- prenos cez informačný kanál
- dekódovanie v dekóderi
- príjem správ u príjemcu

Prv, než sa začneme zaoberať matematickým popisom jednotlivých fáz prenosu informácie, bude treba vyriešiť problém, čomu informáciu priradiť a ako ju kvantifikovať. Tu by mohol neskúsený mladý adept informatiky podľahnúť pokušeniu stotožniť množstvo informácie s veľkosťou súboru, do ktorého sa táto

informácia zapíše. Kto si však spomenie na komprimačné programy (PKZIP, ARJ, RAR,...) zistí, že existuje veľa plnohodnotných spôsobov zápisu, t. j. zakódovania tej istej informácie do súboru, každý z nich vedie k inej veľkosti výsledného súboru. Informáciu teda nie je možné stotožniť s dátami, ktoré predstavujú jej zápis.

Väčšina literatúry o teórii informácie začína uvedením Shannonovej-Hartleyovej formuly $I(A) = -\log_2 P(A)$ bez akejkoľvek motivácie. Čitateľ, ktorý sa dostal k pionierskym prácam o informácii by určite zistil, že cesta k tejto formule nebola priamočiara. V prvej kapitole o zavedení informácie sa snažím ukázať túto motiváciu, kde okrem tradičného prístupu, ktorý priraďuje informáciu javom pravdepodobnostného priestoru, uvádzam i (podľa mňa mimoriadne krásny) spôsob zavedenia informácie bez pojmu pravdepodobnosti podľa Černého a Brunovského z článku [4].

Matematický model informačného zdroja možno formulovať pomocou pojmov elementárnej teórie pravdepodobnosti, ale uvádzam tu i model založený na pojme súčinu spočítateľného počtu priestorov s mierou. Druhý spôsob už predpokladá isté znalosti z teórie miery, avšak umožňuje elegantne definovať niektoré pojmy ako ergodicitu zdroja (a neskôr i ergodicitu kanála). Súčasťou štúdia informačných zdrojov je zavedenie jeho informačnej výdatnosti – *entropie*.

Hlavnou úlohou kódovania správ je prispôsobiť abecedu správy abecede, v ktorej pracuje kanál. Kódovanie má však aj ďalšie úlohy, a to kompresiu správ, schopnosť kódu zistiť, že nastal istý počet chýb pri prenose, alebo dokonca schopnosť opraviť niekoľko chýb pri prenose. Kompresia na jednej strane a schopnosť objavovať, alebo opravovať chyby sú však protichodné požiadavky a vyhovieť im nie je vždy jednoduché. Ukazuje sa, že pri riešení problémov kódovania sa dajú s výhodou využiť výsledky algebry z oblasti konečných grúp, konečných okruhov a telies a lineárnych priestorov nad konečnými telesami. Za vyvrcholenie spojenia poznatkov o kódovaní a o zdroji považujem základnú vetu o kódovaní zdroja, ktorá hovorí, že entropia zdroja je dolnou hranicou strednej dĺžky binárne skomprimovaných správ zo zdroja.

Prenosový kanál sa dá modelovať prostriedkami elementárnej teórie pravdepodobnosti. V tejto práci sa obmedzím na najjednoduchší stacionárny kanál bez pamäte preto, lebo opisuje najbežnejšie prenosové kanály a je relatívne jednoduchý na matematické skúmanie.

Termín kapacita sa v technickej a fyzikálnej praxi používa v dvoch rôznych významoch: v statickom – koľko jednotiek maximálne pojme daný objekt (kapacita pamäte, kapacita autobusu, kapacita posluchárne) a dynamickom – maximálne koľko jednotiek za jednotku času obslúži (resp. spracuje, prepustí, a pod.)

študovaný objekt (kapacita lanoviek v lyžiarskom stredisku, kapacita internetového pripojenia atď.). Kapacita prenosového kanála je kapacitou dynamického typu.

Ak chceme definovať kapacitu prenosového kanála, môžeme postupovať analogicky ako pri definovaní jeho fyzikálnych analógií. Kapacita križovatky je maximum z počtu vozidiel, ktoré ňou prejdú za jednotku času. Kapacita vodovodného potrubia je maximálne množstvo vody, ktoré ním pretečie za jednotku času. Kapacitu informačného kanála je potom celkom prirodzené definovať ako maximálne množstvo informácie, ktoré môže preniesť za jednotku času.

Podobne by sme mohli kapacitu kanála definovať ako maximum počtu bitov, ktorý kanál prenesie za jednotku času. Táto definícia by obstála v prípade bezchybnej práce kanála. Ktorý reálny kanál však dokáže preniesť dáta bez najmenšej chyby? Reálne prenosové cesty sú v prostredí silného priemyselného elektrického rušenia, šumu, statických atmosferických výbojov a mnohých ďalších rušivých vplyvov. V reálnom informačnom kanáli sa informácia pri prenose môže čiastočne zmeniť alebo i stratiť (skúste pod Linuxom príkaz ping). Určiť kapacitu takéhoto kanála so šumom je už značný teoretický i praktický problém.

Predstavme si, že stojíme pred nasledujúcim problémom: Cez daný kanál sa nám nedarí preniesť bez zdržania informácie z nejakého zdroja. Je to skutočne zapríčinené len nízkou kapacitou kanála, alebo je chyba v nesprávnom kódovaní? Tu sa ukáže, že výsledok porovnania entropie zdroja s kapacitou kanála je rozhodujúci pre prenositeľnosť správ bez zdržania cez kanál danej kapacity. Presne je táto skutočnosť formulovaná vo forme priamej a obrátenej Shannonovej vety.

V tejto publikácii podávam základné definície a vety z oboru teórie informácie a kódovania. Pretože táto učebnica je určená pre inžinierov - informatikov, zložitejšie dôkazy vynechávam; prípadných záujemcov odkazujem na príslušnú literatúru. Dôkazy ukončujem znakom ■, časti kapitol, ktoré sú ťažšie a môžu sa preskočiť bez straty kontinuity označujem hviezdičkou pri názve.

Na teórii informácie ma fascinuje, ako sa tu účelne a logicky spájajú výsledky spojitej i diskrétnej, deterministickej i nedeterministickej matematiky, teórie pravdepodobnosti, teórie miery, teórie čísel i algebry do jednej ucelenej zmysluplnej a aplikovateľnej teórie. Prajem čitateľovi, aby pri štúdiu tejto knižky zažil podobné estetické potešenie ako ja pri jej písaní.

Autor.

Obsah

Kapitola 1

Informácia

1.1 Možnosti a spôsoby zavedenia informácie

Ak žiadame informáciu o odchode rýchlika Tatran zo Žiliny do Bratislavy, môžeme ju dostať presne forme nasledujúcej vety: „*Rýchlik Tatran do Bratislavy odchádza zo Žiliny o 20 hodine a 19 minúte.*“ Priateľ, ktorý si nepamätá presne, nám však môže odpovedať nasledovne: „*Neviem presnú minútu, ale určite Tatran do Bratislavy odchádza zo Žiliny medzi 20:00 a 21:00.*“ Ak nás zaujíma, aká je predpoveď teploty na zajtrajší deň, dostaneme ju nasledovne: „*Zajtra sa bude teplota vzduchu pohybovať medzi 18 až 22 stupňami Celsia.*“ Študent informuje svojich rodičov o výsledku skúšky vetou: „*Zo skúšky z algebry som dostal známku B.*“ Alebo len stručne: „*Skúšku z algebry som urobil.*“ Na začiatku futbalového zápasu reportér odhaduje: „*Na stretnutie sa prišlo podívať 5 až 6 tisíc divákov.*“ V priebehu stretnutia, potom, čo dostal presné údaje od usporiadateľov, spresňuje: „*Na zápas prišlo 5764 platiacich divákov.*“

Každý z uvedených výrokov nesie so sebou istú informáciu. Intuitívne cítime, že presná odpoveď o odchode vlaku (20:19) obsahuje viac informácie ako priateľova (medzi 20:00 a 21:00), hoci aj tá druhá je pre nás v čase núdze užitočná. Každý bude tiež súhlasiť, že „*skúška za B*“ nesie viac informácie ako púhe „*urobil som*“. Podobne údaj 5764 platiacich divákov nesie so sebou viac informácie ako odhad 5 až 6 tisíc divákov.

Rýchlik Tatran môže odchádzať o 00:00, 00:01, 00:03, ..., 23:58, 23:59 – existuje 1440 možností odpovede. Pre výsledok skúšky z algebry existuje v našom hodnotiacom systéme 6 možností (A, B, C, D, E, FX). Ľahšie uhádneme

výsledok skúšky z algebry, ako hodinu a minútu odchodu rýchlika. Intuitívne cítime, že v presnej odpovedi na otázku o odchode vlaku je viac informácie ako v odpovedi o výsledku skúšky. Ako však kvantifikovať množstvo informácie?

Dá sa predpokladať, že informácia bude definovaná ako reálna funkcia $I : \mathcal{A} \to \mathbb{R}$ (kde $\mathbb{R}$ je množina reálnych čísel), ktorá každému prvku z nejakej množiny $\mathcal{A}$ priradí nezáporné reálne číslo – množstvo informácie. Prvý problém spočíva v špecifikovaní množiny $\mathcal{A}$. Na prvý pohľad by sa mohlo zdať vhodné brať za množinu $\mathcal{A}$ množinu výrokov. Pracovať s výrokmi však nie je veľmi pohodlné. Radšej by sme mali do činenia s nejakými štandardnejšími matematickými objektami. Každý výrok nesúci informáciu je vlastne veta v tvare: „*Nastal jav A.*“ resp. „*Nastane jav A.*“ **Jav** A v teórii informácie môžeme, podobne ako v teórii pravdepodobnosti, definovať ako podmnožinu množiny Ω všetkých elementárnych (t. j. ďalej už nerozložiteľných) javov. Informáciu teda budeme priraďovať množinám – podmnožinám tzv. základnej množiny Ω elementárnych javov[1].

Ak je Ω nekonečná množina, môžu pri definovaní informácie množiny A vzniknúť isté teoretické problémy súvisiace s merateľnosťou[2] množiny A. Ako neskôr uvidíme, informácia množiny A závisí na jej pravdepodobnostnej miere. Preto sa obmedzíme iba na také systémy podmnožín množiny Ω, pre ktoré vieme určiť ich mieru. Ukazuje sa, že systémy merateľných množín obsahujú samotnú základnú množinu Ω a sú uzavreté na operáciu doplnku a spočítateľného zjednotenia.

Definícia 1.1. Nech Ω je neprázdna množina, ktorú budeme volať aj **základný priestor**. **σ-algebrou** podmnožín základného priestoru Ω nazývame taký systém $\mathcal{A}$ podmnožín množiny Ω, pre ktorý platí:

1. $\Omega \in \mathcal{A}$
2. Ak $A \in \mathcal{A}$ potom aj $A^C = (\Omega - A) \in \mathcal{A}$
3. Ak $A_n \in \mathcal{A}$ pre $n = 1, 2, \ldots,$ potom aj $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

[1] Je výhodné zobrať za Ω jednu univerzálnu množinu obsahujúcu všetky možné elementárne javy (pre celý vesmír a každý čas). Predpokladajme, že pre každý jav $A \subseteq \Omega$ máme orákulum, ktoré pre každé $\omega \in \Omega$ rozhodne, či $\omega \in A$ alebo nie – t. j. funkciu $\chi_A : \Omega \to \{0, 1\}$ takú, že ak $\omega \in A$, potom $\chi_A(\omega) = 1$, ak $\omega \notin A$, potom $\chi_A(\omega) = 0$,

[2] Merateľná množina je taká množina, ktorej možno priradiť Lebesgueovu mieru. Ukazuje sa, že nie všetky podmnožiny množiny reálnych čísel sú merateľné. Čitateľ, ktorý sa zaujíma o praktické aplikácie matematiky, sa však problémom nemerateľnosti nemusí znepokojovať, pretože všetky autorovi známe príklady nemerateľných množín boli zostrojené použitím axiómu výberu – teda nekonštruktívne, a preto je každá praktická množina merateľná.

Vidíme, že σ-algebra obsahuje základný priestor Ω, s každou (aj konečnou) postupnosťou množín obsahuje aj ich zjednotenie, a s každou množinou A obsahuje aj jej doplnok A^C. Dá sa ľahko ukázať, že σ-algebra obsahuje prázdnu množinu $\emptyset$, a s každou (aj konečnou) postupnosťou množín obsahuje aj ich spoločný prienik.

Ak teda vezmeme za množinu $\mathcal{A}$ σ-algebru merateľných podmnožín nejakého základného priestoru Ω, máme prvý problém vyriešený.

Druhým problémom je, ako zaviesť reálnu funkciu $I : \mathcal{A} \to \mathbb{R}$ (kde $\mathbb{R}$ je množina reálnych čísel) tak, aby hodnota $I(A)$ pre $A \in \mathcal{A}$ vyjadrovala informáciu, ktorú dostaneme v správe, že nastal jav A.

V analogickej situácii sme boli, keď sme zavádzali pravdepodobnosť na σ-algebre $\mathcal{A}$, kde sa dalo postupovať trojako – elementárne, axiomaticky a pomocou pojmu normovanej miery na merateľnom priestore $(\Omega, \mathcal{A})$. Pre naše účely bude stačiť analógia elementárneho prístupu. Tento postup zavedenia pravdepodobnosti sa dá charakterizovať nasledovne:

Predpokladáme, že základný priestor Ω je zjednotením konečného počtu n rovnako pravdepodobných disjunktných javov:

$$\Omega = A_1 \cup A_2 \cup \cdots \cup A_n \ ,$$

pravdepodobnosť každého z nich sa musí rovnať $\frac{1}{n}$ – t. j. $P(A_i) = \frac{1}{n}$ pre každé $i = 1, 2, \ldots, n$.
Za prvky σ-algebry $\mathcal{A}$ berieme $\emptyset$ a všetky konečné zjednotenia typu

$$A = \bigcup_{k=1}^{m} A_{i_k}, \tag{1.1}$$

kde $A_{i_k} \neq A_{i_l}$ pre $k \neq l$. Potom každej množine $A \in \mathcal{A}$ tvaru 1.1 priradíme pravdepodobnosť $P(A) = \frac{m}{n}$. Tento postup možno zovšeobecniť aj pre prípad, kedy množiny $A_1, A_2, \ldots, A_n$ majú nerovnaké pravdepodobnosti $p_1, p_2, \ldots, p_n$, kde $p_1 + p_2 + \cdots + p_n = 1$.

Základnou vlastnosťou pravdepodobnosti $P(A)$ na množine $\mathcal{A}$ je aditivita – pre $A, B \in \mathcal{A}$ také, že $A \cap B = \emptyset$ je $P(A \cup B) = P(A) + P(B)$. Pre informáciu $I(A)$ však očakávame, že ak $A \subseteq B$, potom $I(B) \geq I(A)$, t. j. že informácia „menšieho“ javu A je väčšia než informácia „väčšieho“ javu B. Z toho vyplýva, že $I(A \cup B) \leq I(A)$, $I(A \cup B) \leq I(B)$, a preto pre nenulové $I(A)$, $I(B)$ nemôže platiť $I(A \cup B) = I(A) + I(B)$.

Myšlienka ďalšieho postupu je nasledovná. Keďže binárna operácia

$$+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

nevyhovuje pre vyjadrenie informácie disjunktného zjednotenia pomocou informácií zložiek, zavedieme inú operáciu

$$\oplus : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}_0^+,$$

kde $\mathbb{R}_0^+$ je množina nezáporných reálnych čísel, pomocou ktorej vyjadríme informáciu disjunktného zjednotenia ľubovoľných dvoch disjunktných množín A, B nasledovne

$$I(A \cup B) = I(A) \oplus I(B).$$

Pochopiteľne, zatiaľ ešte nevieme, či vôbec taká operácia existuje, a ak existuje, či je takých operácii viac, a ak áno, ako sa od seba líšia. Všimnime si ešte, že na rozdiel od operácie sčítania $+$, ktorá je definovaná na karteziánskom súčine $\mathbb{R} \times \mathbb{R}$, nám stačí, aby bola operácia $\oplus$ definovaná pre nezáporné reálne čísla, t. j. aby bola definovaná na množine $\mathbb{R}_0^+ \times \mathbb{R}_0^+$.

Spíšme si, aké vlastnosti očakávame od informácie

1. $I(A) \geq 0$ pre všetky $A \in \mathcal{A}$ (1.2)
2. $I(\Omega) = 0$ (1.3)
3. Ak $A \in \mathcal{A},\ B \in \mathcal{A},\ A \cap B = \emptyset,$ potom $I(A \cup B) = I(A) \oplus I(B)$ (1.4)
4. Ak $A_n \nearrow A = \bigcup_{i=1}^{\infty} A_i,$ alebo $A_n \searrow A = \bigcap_{i=1}^{\infty} A_i,$ potom $I(A_n) \to I(A)$. (1.5)

Vlastnosť 1. hovorí, že množstvo informácie je nezáporné číslo, vlastnosť 2. hovorí, že zo správy, že nastal jav Ω, nezískame žiadnu informáciu. Vlastnosť 3. hovorí, že informáciu disjunktného zjednotenia javov dostaneme z informácií jednotlivých javov pomocou operácie $\oplus$ a posledná 4. vlastnosť hovorí[3], že informácia je v istom zmysle „spojitá“ na $\mathcal{A}$.

Majme dva javy A, B ktoré so sebou nesú informáciu $I(A)$, $I(B)$. Môže sa stať, že skutočnosť, že nastal jeden z nich, nedáva žiadnu informáciu o druhom. V tom prípade je informácia $I(A \cap B)$ javu $A \cap B$ rovná súčtu informácií oboch javov. Z toho nasledujúca definícia:

Definícia 1.2. Hovoríme, že javy A, B sú **nezávislé**, ak platí

$$I(A \cap B) = I(A) + I(B)\ . \tag{1.6}$$

[3]Zápis $A_n \nearrow A$ hovorí, že $A_1 \subseteq A_2 \subseteq A_3, \ldots$ a $A = \bigcup_{i=1}^{\infty} A_i$. Podobne $A_n \searrow A$ znamená, že $A_1 \supseteq A_2 \supseteq A_3, \ldots$ a $A = \bigcap_{i=1}^{\infty} A_i$. $I(A_n) \to I(A)$ znamená, že $\lim_{n\to\infty} = I(A)$.

Teraz zosumarizujeme vlastnosti binárnej operácie $\oplus$

1. $x \oplus y = y \oplus x$ (1.7)
2. $(x \oplus y) \oplus z \ = x \oplus (y \oplus z)$ (1.8)
3. $I(A) \oplus I(A^C) = 0$ (1.9)
4. $\oplus : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ je spojitá funkcia dvoch premenných (1.10)
5. $(x + z) \oplus (y + z) = (x \oplus y) + z$ (1.11)

Vlastnosti 1. a 2. vyplývajú z komutativity a asociativity množinového zjednotenia. Vlastnosť 3. možno odvodiť z požiadavky $I(\Omega) = 0$ nasledujúcou postupnosťou rovností

$$0 = I(\Omega) = I(A \cup A^C) = I(A) \oplus I(A^C)$$

Vlastnosť 4. – spojitosť je prirodzená požiadavka vyplývajúca zo 4. požiadavky na spojitosť informácie I.

Ostáva objasniť, odkiaľ sa vzala požiadavka 5. Majme dva disjunktné javy A, B také, že A je nezávislé od C a tiež B je nezávislé od C. Ak z toho, že nastal jav A, sa nič nedozviem o jave C a ani z toho, že nastal jav B, sa nič nedozviem o jave C, potom ani z toho, že nastal jav $A \cup B$, nedostanem žiadnu informáciu o jave C, a teda javy $A \cup B$ a jav C sú nezávislé. Označme $x = I(A)$, $y = I(B)$, $z = I(C)$ a počítajme informáciu $I\left[(A \cup B) \cap C)\right]$

$$I\left[(A \cup B) \cap C)\right] = I(A \cup B) + I(C) = I(A) \oplus I(B) + I(C) = x \oplus y + z \quad (1.12)$$

$$\begin{aligned} I\left[(A \cup B) \cap C)\right] &= I\left[(A \cap C) \cup (B) \cap C)\right] = I(A \cap C) \oplus I(B \cap C) = \\ &= \left[I(A) + I(C)\right] \oplus \left[I(B) + I(C)\right] = (x + z) \oplus (y + z) \end{aligned} \quad (1.13)$$

Porovnaním vzťahov (1.12), (1.13) dostaneme žiadanú vlastnosť 5.

Veta 1.1. *Nech binárna operácia $\oplus$ na množine $\mathbb{R}_0^+$ vyhovuje axiómam* (1.7) *až* (1.11). *Potom*

buď $\quad \forall x, y \in \mathbb{R}_0^+ \qquad x \oplus y = \min\{x, y\},$ (1.14)

alebo $\quad \exists k > 0 \ \forall x, y \in \mathbb{R}_0^+ \qquad x \oplus y = -k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right).$ (1.15)

Dôkaz tejto vety je zložitý, čitateľ ho nájde v [5].

Zaujímavé ale je, že (1.14) je limitným prípadom (1.15) pre $k \to 0+$. Nech najprv $x = y$ a teda $\min\{x, y\} = x$. Potom

$$-k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right) = -k \log_2 \left(2.2^{-\frac{x}{k}}\right) =$$
$$= -k \log_2 \left(2^{(-\frac{x}{k}+1)}\right) = -k.(-\frac{x}{k} + 1) = x - k$$

Teraz je už vidieť, že predchádzajúci výraz konverguje k x pre $k \to 0^+$. Nech $x > y$, potom $\min\{x, y\} = y$. Platí:

$$-k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right) = -k \log_2 \left(2^{-\frac{y}{k}}.(2^{\frac{y-x}{k}} + 1)\right) = y - k. \log_2 \left(2^{\frac{y-x}{k}} + 1\right)$$

Na dokázanie nášho tvrdenia stačí ukázať, že druhý člen posledného rozdielu konverguje k 0 ak $k \to 0^+$. Použitím l'Hospitalovho pravidla máme

$$\lim_{k\to 0^+} k. \log_2 \left(2^{\frac{y-x}{k}} + 1\right) = \lim_{k\to 0^+} \frac{\log_2 \left(2^{\frac{y-x}{k}} + 1\right)}{\frac{1}{k}} =$$
$$= \lim_{k\to 0^+} \frac{\frac{2^{(y-x)/k}.\ln(2).(y-x)}{(2^{(y-x)/k}+1)/k^2}}{\frac{1}{k^2}} = \ln(2)(y - x). \lim_{k\to 0^+} \frac{2^{(y-x)/k}}{2^{(y-x)/k} + 1} = 0$$

pretože $(y - x) < 0$, $(y - x)/k \to -\infty$ pre $k \to 0^+$, a preto $2^{(y-x)/k} \to 0$. Je teda $\lim_{k\to 0^+} -k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right) = \min\{x, y\}$.

Veta 1.2. *Nech* $x \oplus y = -k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right)$ *pre všetky nezáporné reálne* x, y. *Nech* $x_1, x_2, \ldots, x_n$ *sú nezáporné reálne čísla. Potom*

$$\bigoplus_{i=1}^{n} x_i = x_1 \oplus x_2 \oplus \cdots \oplus x_n = -k \log_2 \left(2^{-\frac{x_1}{k}} + 2^{-\frac{x_2}{k}} + \cdots + 2^{-\frac{x_n}{k}}\right) \quad (1.16)$$

Dôkaz. Dôkaz matematickou indukciou podľa n si čitateľ ľahko urobí sám. ■

1.2 Elementárna definícia informácie

Keď už máme definovanú operáciu $\oplus$, môžeme sa pokúsiť zaviesť množstvo informácie analogicky ako to robí elementárna definícia pravdepodobnosti. Nech $\{A_1, A_2, \ldots, A_n\}$ je rozklad priestoru Ω na javy s rovnakou informáciou, t. j. nech

$$1. \quad \Omega = \bigcup_{i=1}^{n} A_i, \text{ kde } A_i \cap A_j = \emptyset \text{ pre } i \neq j \tag{1.17}$$

$$2. \quad I(A_1) = I(A_2) = \cdots = I(A_n) = a \text{ pre } i \neq j \tag{1.18}$$

Chceme určiť veličinu a. Z (1.17), (1.18) vyplýva

$$0 = I(\Omega) = I(A_1) \oplus I(A_2) \oplus \cdots \oplus I(A_n) = \underbrace{a \oplus a \oplus \cdots \oplus a}_{n-\text{krát}} = \bigoplus_{i=1}^{n} a \tag{1.19}$$

$$0 = \bigoplus_{i=1}^{n} a = \\ = \begin{cases} \min\{a, a, \ldots, a\} = a & \text{ak } x \oplus y = \min\{x, y\} \\ -k \log_2 \left(2^{-a/k} + \cdots + 2^{-a/k}\right) & \text{ak } x \oplus y = -k \log_2 \left(2^{-x/k} + 2^{-y/k}\right) \end{cases} \tag{1.20}$$

Pre prvý prípad $\bigoplus_{i=1}^{n} = a = 0$ a teda každý jav rozkladu $\{A_1, A_2, \ldots, A_n\}$ nesie so sebou nulovú informáciu. Toto je výsledok nezaujímavý a nemá význam sa ním ďalej zaoberať.

Pre druhý prípad

$$\bigoplus_{i=1}^{n} a = -k \log_2 \left(\underbrace{2^{-a/k} + \cdots + 2^{-a/k}}_{n-\text{krát}} \right) = -k \log_2 \left(n.2^{-a/k}\right) = a - k \log_2(n) = 0$$

Z posledného vzťahu vyplýva, že

$$a = k. \log_2(n) = -k. \log_2 \left(\frac{1}{n}\right) \tag{1.21}$$

Nech jav A je zjednotením m rôznych základných javov $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$. Potom

$$\begin{aligned} I(A) &= I(A_{i_1}) \oplus I(A_{i_2}) \oplus \cdots \oplus I(A_{i_m}) = \underbrace{a \oplus a \oplus \cdots \oplus a}_{m\text{–krát}} = \\ &= -k.\log_2 \left(\underbrace{2^{-a/k} + 2^{-a/k} + \cdots + 2^{-a/k}}_{m\text{–krát}} \right) = -k \log_2 \left(m.2^{-a/k} \right) = \\ &= -k.\log_2(m) - k.\log_2 \left(2^{-a/k} \right) = -k.\log_2(m) - k.(-a/k) = \\ &= -k.\log_2(m) + a \ = -k.\log_2(m) + k.\log_2(n) = \\ &= k.\log_2 \left(\frac{n}{m} \right) = -k.\log_2 \left(\frac{m}{n} \right) \end{aligned} \tag{1.22}$$

Veta 1.3. *Nech* $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$ *je rozklad priestoru* Ω *na javy s rovnakou informáciou. Potom pre informáciu* $I(A_i)$ *každého javu* A_i $i = 1, 2, \ldots, n$ *platí*

$$I(A_i) = -k \log_2 \frac{1}{n}. \tag{1.23}$$

Nech $A = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_m}$ *je zjednotenie* m *rôznych množín rozkladu* $\mathbf{A}$, *t. j.* $A_{i_k} \in \mathbf{A}$, $A_{i_k} \neq A_{i_l}$ *pre* $k \neq l$. *Potom pre informáciu* $I(A)$ *javu* A *platí*

$$I(A) = -k \log_2 \frac{m}{n}. \tag{1.24}$$

Všimnime si zaujímavú analógiu s elementárnym zavedením pravdepodobnosti. Ak je základný priestor Ω rozdelený na n disjunktných javov $A_1, A_2, \ldots, A_n$ s rovnakou pravdepodobnosťou p, potom túto pravdepodobnosť vypočítame zo vzťahu $\sum_{i=1}^{n} p = n.p = 1$ a teda $P(A_i) = p = 1/n$. Ak je nejaká množina A disjunktným zjednotením m množín rozkladu, potom jej pravdepodobnosť vypočítame ako $P(A) = m/n$.

Pri zavádzaní informácie sa informácia a každej množiny A_i rozkladu určí pomocou vzťahu (1.20), odkiaľ máme $I(A_i) = a = -k.\log_2(1/n)$ a informácia množiny A, ktorá je zjednotením m disjunktných množín rozkladu sa vypočíta ako $I(A) = -k.\log_2(m/n)$.

Zastavme sa ešte na chvíľu pri konštante k. Táto závisí od toho, ako zvolíme jednotku informácie. Rôznym hodnotám k odpovedá rôzna mierka určovania veľkosti informácie. (Pri číselnom vyjadrovaní vzdialenosti tiež výsledok závisí od toho, či túto vzdialenosť vyjadrujeme v kilometroch, míľach či yardoch.)

Pre prechod od sústavy logaritmov so základom a k logaritmom so základom b platí známy vzorec

$$\log_b(x) = \log_b(a).\log_a(x) = \frac{1}{\log_a(b)}.\log_a(x). \tag{1.25}$$

Namiesto konštanty k a logaritmu pri základe 2 by mohol vo vzťahoch (1.21), (1.22) vystupovať iba logaritmus pri ľubovoľnom základe. Toto skutočne niektorí autori aj používajú, najmä v staršej literatúre sa občas objaví používanie dekadického logaritmu.

Pre určenie konštanty k môže byť užitočná nasledujúca úvaha. Výpočtová a digitálna prenosová technika prenáša informáciu prevažne pomocou binárnych znakov, ktoré môžu nadobúdať len dve hodnoty 0 a 1. Bolo by prirodzené, keby jeden takýto znak prenášal jednotkové množstvo informácie, ktoré nazveme 1 bit.

Nech $\Omega = \{0, 1\}$ je množina hodnôt, ktoré môže nadobúdať jeden binárny znak, $A_1 = \{0\}$, $A_2 = \{1\}$, nech obe tieto jednoprvkové množiny nesú rovnakú informáciu a, ktorú by sme radi prehlásili za jednotkovú. Chceme, aby $I(A_1) = I(A_2) = a = 1$. Podľa (1.21) $1 = a = k.\log_2(2) = k$. Ak teda chceme, aby vzorce (1.21), (1.22) vyjadrovali množstvo informácie v bitoch, musíme v nich položiť $k = 1$. Odteraz budeme predpokladať, že informáciu meriame v bitoch a teda že k=1.

1.3 Informácia ako funkcia pravdepodobnosti

Pri elementárnej definícii informácie na základe rozkladu $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$ sme odvodili, že veľkosť informácie pre množinu A, ktorá je zjednotením m základných množín je $I(A) = -\log_2(m/n)$, pravdepodobnosť množiny A je $P(A) = m/n$. V tomto prípade by sa dalo písať $I(A) = -\log_2(P(A))$.

Pokúsme sa dostať k definícii informácie z iného konca, a to pomocou pravdepodobnosti. Predpokladajme teda, že informácia $I(A)$ javu A závisí iba od pravdepodobnosti $P(A)$ javu A, t. j. $I(A) = f(P(A))$, pričom funkcia f nezávisí od toho, aký je pravdepodobnostný priestor $(\Omega, \mathcal{A}, P)$. Aké možné funkcie pripadajú do úvahy na mieste funkcie f? Ukážeme, že jedinou funkciou pripadajúcou do úvahy je funkcia $f(x) = -k.\log_2(x)$. Použijeme pritom postup podľa [5].

Najprv však rozšírime definíciu 1.2 nezávislej dvojice javov nasledovne.

Definícia 1.3. Konečná alebo nekonečná postupnosť javov $\{A_n\}_n$ sa nazýva **postupnosťou (informačne) nezávislých javov**, ak pre každú konečnú vybranú postupnosť $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ platí

$$I\left(\bigcap_{k=1}^{m} A_{i_k}\right) = \sum_{k=1}^{m} I\left(A_{i_k}\right). \tag{1.26}$$

Aby informácia mala „rozumné“ vlastnosti, treba požadovať, aby funkcia f bola spojitá a aby javy nezávislé v pravdepodobnostnom zmysle ostali nezávislými v zmysle teórie informácie. To znamená, že pre postupnosť nezávislých javov $A_1, A_2, \ldots, A_n$ platí

$$I(A_1 \cap A_2 \cap \cdots \cap A_n) = f(P(A_1 \cap A_2 \cap \cdots \cap A_n)) = f\left(\prod_{i-1}^{n} P(A_i)\right) \tag{1.27}$$

a súčasne

$$I(A_1 \cap A_2 \cap \cdots \cap A_n) = \sum_{i=1}^{n} I(A_i) = \sum_{i-1}^{n} f\left(P(A_i)\right) \tag{1.28}$$

Ľavé strany posledných dvoch vzťahov musia byť rovnaké, a preto

$$f\left(\prod_{i-1}^{n} P(A_i)\right) = \sum_{i-1}^{n} f\left(P(A_i)\right) \tag{1.29}$$

Nech sú všetky javy $A_1, A_2, \ldots, A_n$ rovnako pravdepodobné, nech $P(A_i) = x$. Potom $f(x^n) = n.f(x)$ pre všetky $x \in \langle 0, 1\rangle$. Pre $x = 1/2$ máme

$$f(x^m) = f\left(\frac{1}{2^m}\right) = m.f\left(\frac{1}{2}\right). \tag{1.30}$$

Pre $x = \dfrac{1}{2^{1/n}}$ je $f(x^n) = f\left(\left(\dfrac{1}{2^{1/n}}\right)^n\right) = f\left(\dfrac{1}{2}\right) = n.f(x) = n.f\left(\dfrac{1}{2^{1/n}}\right)$,
z čoho máme

$$f\left(\frac{1}{2^{1/n}}\right) = \frac{1}{n}.f\left(\frac{1}{2}\right) \tag{1.31}$$

Konečne pre $x = \dfrac{1}{2^{1/n}}$ je

$$f(x^m) = f\left(\frac{1}{2^{m/n}}\right) = m.f(x) = m.f\left(\frac{1}{2^{1/n}}\right) = \frac{m}{n}.f\left(\frac{1}{2}\right),$$

takže

$$f\left(\frac{1}{2^{m/n}}\right) = \frac{m}{n}.f\left(\frac{1}{2}\right) \tag{1.32}$$

Pretože (1.32) platí pre všetky kladné čísla m, n a pretože predpokladáme, že funkcia f je spojitá, musí platiť

$$f\left(\frac{1}{2^x}\right) = x.f\left(\frac{1}{2}\right) \text{ pre všetky reálne čísla } x \in \langle 0, \infty).$$

Vytvorme pomocnú funkciu g predpisom $g(x) = f(x) + f\left(\frac{1}{2}\right).\log_2(x)$.
Potom platí

$$\begin{aligned} g(x) &= f(x) + f\left(\frac{1}{2}\right).\log_2(x) = f\left(2^{\log_2(x)}\right) + f\left(\frac{1}{2}\right).\log_2(x) = \\ &= f\left(\frac{1}{2^{-\log_2(x)}}\right) + f\left(\frac{1}{2}\right).\log_2(x) = \\ &= -\log_2(x).f\left(\frac{1}{2}\right) + f\left(\frac{1}{2}\right).\log_2(x) = 0 \end{aligned}$$

Funkcia $g(x) = f(x) + f\left(\frac{1}{2}\right).\log_2(x)$ je identicky 0, a preto

$$f(x) = -f\left(\frac{1}{2}\right).\log_2(x) = -k.\log_2(x) \tag{1.33}$$

Pre množstvo informácie z posledného vzťahu vyplýva slávna **Shannonova – Hartleyova formula:**

$$I(A) = -k.\log_2(P(A)) \tag{1.34}$$

Podobne, ako pri elementárnom spôsobe definovania množstva informácie, aj tu vystupuje koeficient k závislý na mierke určovania veľkosti informácie.

Nech $\Omega = \{0, 1\}$ je množina hodnôt, ktoré môže nadobúdať jeden binárny znak, $A_1 = \{0\}$, $A_2 = \{1\}$, nech obe tieto jednoprvkové množiny majú rovnakú pravdepodobnosť $P(A_1) = P(A_2) = 1/2$. Keďže veľkosť informácie je funkciou pravdepodobnosti, nesú obe množiny A_1, A_2 rovnakú informáciu a, ktorú by sme radi prehlásili za jednotkovú. Preto musí byť $1 = f\left(\frac{1}{2}\right) = -k.\log_2\left(\frac{1}{2}\right) = k$,

čiže $k = 1$. Aj pri tomto prístupe sme došli k analogickému výsledku ako pri elementárnej definícii informácie.

V mnohých učebniciach je čitateľ postavený pred hotovú definíciu veľkosti informácie pomocou Shannonovej-Hartleyovej formuly $I(A) = -\log_2 P(A)$, z ktorej sa potom odvádzajú jej vlastnosti. Čitateľ sa môže spýtať, prečo je veľkosť informácie definovaná práve takouto formulou. Elementárny, axiomatický i pravdepodobnostný spôsob zavedenia informácie ukazujú, že je to tak preto, lebo sú na to rozumné dôvody a ináč sa to proste nedá.

Kapitola 2

Entropia

2.1 Pokusy

Ak dostaneme správu, že nastal jav $A \in \mathcal{A}$ s pravdepodobnosťou $P(A)$, dostaneme s ňou informáciu $-\log_2 P(A)$ bitov. Predstavme si teraz, že máme základnú množinu javov Ω rozdelenú na konečný počet disjunktných javov $A_1, A_2, \ldots, A_n$. Chceme uskutočniť pokus na určenie toho javu A_i, ktorý nastal.

Pred vykonaním pokusu máme neistotu o jeho výsledku. Po uskutočnení pokusu sa výsledok dozvieme a naša neistota zmizne. Môžeme teda povedať, že veľkosť neistoty pred pokusom sa rovná množstvu informácie, ktorú nám dodá vykonanie pokusu.

V niektorých prípadoch môžeme pokus organizovať – môžeme určiť, aké budú jednotlivé množiny rozkladu, čo chceme urobiť tak, aby sme dostali po vykonaní pokusu čo najväčšiu informáciu. Rozklad množiny Ω na javy, z ktorých každý zodpovedá jednému z výsledkov pokusu, volíme podľa vhodne zvolenej otázky, súboru otázok, možností meracieho postupu a podobne. Správne zvolený experiment je v mnohých odboroch ľudskej činnosti jedným z rozhodujúcich predpokladov úspechu.

Pokúsme sa teraz matematicky vyjadriť to, o čom hovorili predošlé tri odstavce.

Definícia 2.1. Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. **Konečný merateľný rozklad istého javu** je konečná množina javov (t. j. podmnožín Ω) $\{A_1, A_2, \ldots, A_n\}$ taká, že $A_i \in \mathcal{A}$ pre $i = 1, 2, \ldots, n$, $\bigcup_{i=1}^{n} A_i = \Omega$ a $A_i \cap A_j = \emptyset$

pre $i \neq j$. Konečný merateľný rozklad $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$ istého javu Ω nazývame tiež **pokusom**.

V niektorej literatúre sa od množín $\{A_1, A_2, \ldots, A_n\}$ pokusu $\mathbf{P}$ žiadajú oslabené predpoklady, a to $P\left(\bigcup_{i=1}^{n} A_i\right) = 1$ a $P(A_i \cap A_j) = 0$ pre $i \neq j$. Oba prístupy sú prakticky rovnocenné a výsledky jedného sa dajú preniesť na výsledky druhého.

Pokus by mal byť organizovaný tak, aby jeho vykonanie prinášalo čo najväčšiu informáciu. Ak chcem zistiť odchod vlaku, viac informácie mi dá otázka „O ktorej hodine a minúte odchádza vlak Tatran zo Žiliny?“ ako otázka „Odchádza vlak Tatran zo Žiliny predpoludním alebo popoludní?“. Prvá otázka rozdeľuje priestor Ω možných výsledkov na 1440 javov, druhá otázka len na dva javy.

Obe otázky teda definujú dva pokusy $\mathbf{P}_1$, $\mathbf{P}_2$. Za predpokladu, že všetky javy sú v rámci svojho pokusu rovnako pravdepodobné, majú všetky javy pokusu $\mathbf{P}_1$ pravdepodobnosť $1/1440$, javy pokusu $\mathbf{P}_2$ majú pravdepodobnosť $1/2$. Ktorýkoľvek jav pokusu $\mathbf{P}_1$ prináša $-\log_2(1/1440) = 10.49$ bitov, oba javy pokusu $\mathbf{P}_2$ prinášajú rovnakú informáciu $-\log_2(1/2) = 1$ bit.

Nezávisle na tom, aký bol výsledok pokusu $\mathbf{P}_1$ dostaneme informáciu 10.49 bitov, podobne po vykonaní pokusu $\mathbf{P}_2$ dostaneme vždy informáciu 1 bit. Množstvo informácie, ktoré dostaneme po vykonaní takéhoto pokusu, budeme považovať tiež za mieru jeho neurčitosti pred jeho vykonaním – entropiu pokusu.

2.2 Shannonova definícia entropie

Vieme teda definovať neurčitosť - entropiu $H(\mathbf{P})$ pokusu $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$, ak všetky jeho javy A_i majú rovnakú pravdepodobnosť $1/n$ – v tomto prípade je $H(\mathbf{P}) = -\log_2(1/n)$.

Čo však v prípade, keď javy pokusu nemajú rovnakú pravdepodobnosť? Predstavme si, že $\Omega = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, $P(A_1) = 0.1$, $P(A_2) = 0.9$. Ak výsledkom pokusu bude A_1, dostaneme informáciu $I(A_1) = -\log_2(0.1) = 3.32$ bitov, ale ak vyjde A_2, dostaneme informáciu len $I(A_2) = -\log_2(0.9) = 0.15$ bitu. Výsledná informácia teda závisí na výsledku pokusu. Ak vyjde A_1, sme na tom výborne, lenže to sa stane len v jednej desatine prípadov. V 90% prípadov však vyjde A_2 a v tejto väčšine prípadov sme na tom so získanou informáciou zle.

Predstavme si teraz, že pokus vykonáme veľký počet krát – napr. 100 krát. Približne v desiatich prípadoch dostaneme informáciu 3.32 bitov, približne v 90 prípadoch dostaneme informáciu 0.15 bitu, celkovú získanú informáciu možno vyčísliť ako $10 \times 3.32 + 90 \times 0.15 = 33.2 + 13.5 = 46.7$ bitov. Priemerná informácia

na jeden pokus je $46.7/100 = 0.467$ bitov. Možnosťou, ako vo všeobecnom prípade zaviesť entropiu pokusu je definovať ju ako strednú hodnotu informácie.

Definícia 2.2. Shannonova definícia entropie. Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor, na ktorom je daná informácia $I(A) = -\log_2 P(A)$. Nech $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$ je pokus. **Entropia $H(\mathbf{P})$ pokusu $\mathbf{P}$** je stredná hodnota diskrétnej náhodnej veličiny X, ktorá nadobúda na podmnožine A_i hodnotu $I(A_i)$,[1] t. j.

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i)P(A_i) = -\sum_{i=1}^{n} P(A_i).\log_2 P(A_i) \tag{2.1}$$

Dôsledný čitateľ by sa teraz mohol spýtať, čo sa stane, keď sa v pokuse $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$ vyskytne množina A_i s nulovou pravdepodobnosťou. Potom totiž výraz $-P(a_i).\log_2 P(A_i)$ nie je definovaný. Pretože $\lim_{x\to 0+} x\log_2(x) = 0$, je prirodzené definovať funkciu $\eta(x)$ nasledovne

$$\eta(x) = \begin{cases} -x.\log_2(x) & \text{ak } x > 0 \\ 0 & \text{ak } x = 0. \end{cases}$$

Potom by Shannonova formula pre entropiu mala byť v tvare

$$H(\mathbf{P}) = \sum_{i=1}^{n} \eta(P(A_i)).$$

Takýto zápis však trochu zastiera tvar nenulových sčítancov formuly, a preto radšej ostaneme pri tvare (2.1) s tým, že učiníme nasledujúcu dohodu:

Dohoda 2.1. Odteraz budeme predpokladať, že výraz $0.\log_2(0)$ je definovaný a že

$$0.\log_2(0) = 0.$$

Toto dobre vyjadruje skutočnosť, že ak k nejakému pokusu $\mathbf{P}$, (t. j. merateľnému rozkladu množiny Ω) pridáme prázdnu množinu – (t. j. nemožný výsledok), dostaneme nový pokus $\mathbf{P}'$, ktorého neurčitosť bude rovnaká, ako pri pokuse $\mathbf{P}$.

[1] Presne by sme mohli definovať náhodnú veličinu X ako

$$X(\omega) = -\sum_{i=1}^{n} \chi_{A_i}(\omega).\log_2 P(A_i).,$$

kde $\chi_{A_i}(\omega)$ je indikátor množiny A_i, t. j. $\chi_{A_i}(\omega) = 1$ práve vtedy, keď $\omega \in A_i$, inak $\chi_{A_i}(\omega) = 0$.

2.3 Axiomatická definícia entropie

Postup pri odvodení Shannonovej formuly v predchádzajúcej časti je jednoduchý a názorný, no nie všetci autori sú s ním spokojní. Nespokojní sú najmä tí, ktorí by radi zaviedli entropiu bez individuálnej informácie $I(A)$ javu A. Skúsme sledovať myšlienkový postup pri zavádzaní miery neurčitosti $H(\mathbf{P})$ pokusu $\mathbf{P}$ bez využitia pojmu informácie.

Majme pokus $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$, nech $p_1 = P(A_1)$, $p_2 = P(A_2)$, $\ldots$, $p_n = P(A_n)$. Predpokladáme, že funkcia H nezávisí od konkrétneho tvaru pravdepodobnostného priestoru $(\Omega, \mathcal{A}, P)$, ale závisí iba od čísel $p_1, p_2, \ldots, p_n$, teda

$$H(\mathbf{P}) = H(p_1, p_2, \ldots, p_n)$$

Funkcia $H(p_1, p_2, \ldots, p_n)$ by mala mať niektoré prirodzené vlastnosti vyplývajúce z jej významu. Tieto vlastnosti možno formulovať ako axiómy, z ktorých potom možno odvodiť vlastnosti, resp. tvar funkcie H.

Existuje niekoľko sústav axióm, my uvedieme tzv. Fadejevovu sústavu z roku 1956:

AF0: Funkcia $y = H(p_1, p_2, \ldots, p_n)$ je definovaná pre všetky n a pre všetky $p_1 \geq 0,\ p_2 \geq 0, \ldots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, a nadobúda reálne hodnoty.

AF1: $y = H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1 \rangle$.

AF2: $y = H(p_1, p_2, \ldots, p_n)$ je symetrická funkcia, t. j. pre ľubovoľnú permutáciu π čísel $1, 2, \ldots, n$ platí:

$$H(p_{\pi[1]}, p_{\pi[2]}, \ldots, p_{\pi[n]}) = H(p_1, p_2, \ldots, p_n) \tag{2.2}$$

AF3: Ak $p_n = q_1 + q_2 > 0$, $q_1 \geq 0$, $q_2 \geq 0$, potom

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2) = \\ = H(p_1, p_2, \ldots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) \tag{2.3}$$

K týmto axiómam pridáme ešte Shannonovu axiómu. Označme

$$F(n) = H\left(\underbrace{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}}_{n\text{–krát}}\right) \tag{2.4}$$

Shannonova axióma znie:

AS4: Ak $m < n$, potom $F(m) < F(n)$.

Axióma A0 je prirodzená – chceme, aby entropia existovala pre všetky možné pokusy a aby bola reálnym číslom. Axióma A1 vyjadruje prirodzenú požiadavku, aby sa pri malej zmene pravdepodobností dvojprvkového pokusu málo zmenila jeho neurčitosť. Axióma A2 hovorí, že nezáleží na poradí, v akom sú vymenované javy pokusu, čo je zas veľmi prirodzená požiadavka.

Na dlhšie sa treba pristaviť pri axióme A3. Predpokladajme, že máme pokus $\mathbf{P} = \{A_1, A_2, \ldots, A_{n-1}, A_n\}$ s pravdepodobnosťami $p_1, p_2, \ldots, p_n$, od ktorého prejdeme k pokusu $\mathbf{P}' = \{A_1, A_2, \ldots, A_{n-1}, B_1, B_2\}$, ktorý vznikne tak, že posledný jav A_n pokusu $\mathbf{P}$ rozdelíme na dve disjunktné časti B_1, B_2. Potom $P(B_1) + P(B_2) = P(A_n)$. Ak označíme $P(B_1) = q_1$, $P(B_2) = q_2$, potom $p_n = q_1 + q_2$. Aký bude prírastok neistoty, ak prejdeme od pokusu $\mathbf{P}$ k pokusu $\mathbf{P}'$?

Ak už nastane jav A_n, tak pri pokuse $\mathbf{P}'$ máme ešte dodatočnú neistotu, či nastal jav B_1 alebo B_2. Podmienené pravdepodobnosti javov B_1, B_2 za predpokladu, že nastal jav A_n, sú $P(B_1 \cap A_n)/P(A_n) = P(B_1)/P(A_n) = q_1/p_n$, $P(B_2 \cap A_n)/P(A_n) = P(B_2)/P(A_n) = q_2/p_n$, takže ak už nastal jav A_n, ostáva nám ešte neurčitosť

$$H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Avšak jav A_n nenastane vždy, ale len s pravdepodobnosťou p_n. Preto rozdelenie množiny A_n na disjunktné javy B_1, B_2 prispeje k celkovej neurčitosti čiastkou

$$p_n.H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Fadejevove axiómy AF0, až AF3 sú dostatočné na odvodenie tvaru funkcie H a dá sa z nich dokázať i platnosť Shannonovej axiómy AS4. Dôkaz je však dosť zložitý, a preto si pre naše účely dodáme celkom prirodzenú Shannonovu axiómu AS4, ktorá vraví, že ak máme dva pokusy $\mathbf{P}_1$, $\mathbf{P}_2$, prvý s m rovnako

pravdepodobnými javmi, druhý s n rovnako pravdepodobnými javmi a $m < n$, potom neurčitosť pokusu $\mathbf{P}_1$ je menšia ako neurčitosť pokusu $\mathbf{P}_2$. Rozpaky pri predvídaní výsledku pokusu s menším počtom rovnocenných javov sú menšie ako rozpaky pri očakávaní výsledku pokusu s väčším počtom rovnocenných javov. Táto požiadavka sa zdá byť veľmi prirodzená a my ju prijmeme ako axiómu.

Veta 2.1. *Shannonova entropia*

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i)P(A_i) = -\sum_{i=1}^{n} P(A_i)\log_2 P(A_i)$$

spĺňa axiómy AF1 až AF3 a Shannonovu axiómu AS4.

Dôkaz. Overenie jednotlivých axióm je jednoduché a priamočiare, čitateľ si si ho ľahko urobí sám. ■

Teraz na základe axióm AF1 až AF3, AS4 dokážeme niekoľko tvrdení, ktoré nám ukážu niektoré zaujímavé vlastnosti funkcie H spĺňajúcej tieto axiómy. Jednotlivé tvrdenia nás postupne dovedú až k Shannonovej entropickej formule. Pretože podľa vety 2.1 Shannonova entropia daná vzorcom (2.1) spĺňa všetky axiómy, nasledujúce vety platia aj pre ňu.

Veta 2.2. *Funkcia $y = H(p_1, p_2, \ldots, p_n)$ je spojitá funkcia na množine*

$$\mathcal{Q}_n = \left\{(x_1, x_2, \ldots, x_n) \mid x_i \geq 0 \text{ pre } i = 1, 2 \ldots, n, \sum_{i=1}^{n} x_i = 1\right\}.$$

Dôkaz. Matematickou indukciou podľa m. Pre $m = 2$ je tvrdenie axiómou AF1. Nech funkcia $y = H(x_1, x_2, \ldots x_m)$ je spojitá na $\mathcal{Q}_m$.
Nech $(p_1, p_2, \ldots, p_m, p_{m+1}) \in \mathcal{Q}_{m+1}$. Predpokladajme, že aspoň jedno z čísel $p_m,\ p_{m+1}$ je nenulové (inak zmeníme poradie čísel p_i). Použitím axiómy AF3 máme:

$$H(p_1, p_2, \ldots, p_m, p_{m+1}) = H\big(p_1, p_2, \ldots, p_{m-1}, (p_m + p_{m+1})\big) + \\ + (p_m + p_{m+1}).H\left(\frac{p_m}{(p_m + p_{m+1})}, \frac{p_{m+1}}{(p_m + p_{m+1})}\right) \quad (2.5)$$

Spojitosť prvého sčítanca pravej strany (2.5) vyplýva z indukčného predpokladu, spojitosť druhého sčítanca vyplýva z axiómy AF1. ■

Veta 2.3. $H(1, 0) = 0$.

Dôkaz.

$$H\left(\frac{1}{2},\underbrace{\frac{1}{2},0}\right) = H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{2}.H(1,0) \tag{2.6}$$

$$H\left(\frac{1}{2},\frac{1}{2},0\right) = H\left(0,\underbrace{\frac{1}{2},\frac{1}{2}}\right) = H(0,1) + H\left(\frac{1}{2},\frac{1}{2}\right) = H\left(\frac{1}{2},\frac{1}{2}\right) + H(1,0) \tag{2.7}$$

Porovnaním pravých strán (2.6), (2.7) dostávame $\frac{1}{2}.H(1,0) = H(1,0)$, z čoho vyplýva $H(1,0) = 0$. ■

Veta (2.3) hovorí, že neurčitosť pokusu pozostávajúceho z dvoch javov, z ktorých je jeden istý druhý nemožný, je nulová.

Veta 2.4. $H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n)$

Dôkaz. Aspoň jedno z čísel $p_1, p_2, \ldots, p_n$ je kladné. Nech je to p_n (inak zmeníme poradie). Potom

$$H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n) + p_n.\underbrace{H(1,0)}_{0} \tag{2.8}$$

■

Zase jedna dobrá vlastnosť entropie – nezávisí na tom, koľko javov s nulovou pravdepodobnosťou sa vyskytuje v rozklade.

Veta 2.5. *Nech* $p_n = q_1 + q_2 + \cdots + q_m > 0$. *Potom*

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) = \\ = H(p_1, p_2, \ldots, p_n) + p_n.H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right) \tag{2.9}$$

Dôkaz. Matematickou indukciou podľa m. Pre $m = 2$ je tvrdenie axiómou AF3.
Nech tvrdenie platí pre nejaké $m \geq 2$.
Položme $p' = q_2 + q_3 + \cdots + q_{m+1}$, predpokladajme, že $p' > 0$ (inak zameníme poradie $q_1, q_2, \ldots, q_{m+1}$). Podľa indukčného predpokladu

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, \underbrace{q_2, \ldots, q_{m+1}}_{p' = \sum_{k=2}^{m} q_k}) =$$

$$= H(p_1, p_2, \ldots, p_{n-1}, \underbrace{q_1, p'}_{p_n}) + p'.H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right) =$$

$$= H(p_1, p_2, \ldots, p_n) + p_n.\left[H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n}H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right)\right]. \quad (2.10)$$

Ďalej podľa indukčného predpokladu platí

$$H\left(\frac{q_1}{p_n}, \underbrace{\frac{q_2}{p_n}, \ldots, \frac{q_{m+1}}{p_n}}_{\frac{p'}{p_n}}\right) = H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n}H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right). \quad (2.11)$$

Vidíme, že pravá strana (2.11) je totožná s obsahom veľkej hranatej zátvorky na pravej strane vzťahu (2.10). Dosadením ľavej strany vzťahu (2.11) do (2.10) dostávame (2.9). ■

Veta 2.6. *Nech pre* $i = 1, 2, \ldots, n$ *máme* $p_i = q_{i1} + q_{i2} + \cdots + q_{im_i} > 0$. *Potom*

$$H(q_{11}, q_{12} \ldots q_{1m_1}, q_{21}, q_{22}, \ldots, q_{2m_2}, \ldots, q_{n1}, q_{n2}, \ldots, q_{nm_n}) =$$
$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i.H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \ldots, \frac{q_{im_i}}{p_i}\right) \quad (2.12)$$

Dôkaz. Opakovaným použitím vety (2.5). ■

Veta 2.7. *Označme* $F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$. *Potom* $F(mn) = F(m) + F(n)$.

Dôkaz. Použitím vety 2.6 máme

$$\begin{aligned} F(mn) &= H\Bigg(\underbrace{\underbrace{\frac{1}{mn}, \ldots, \frac{1}{mn}}_{m\text{-krát}}, \ldots \underbrace{\frac{1}{mn}, \ldots, \frac{1}{mn}}_{m\text{-krát}}}_{n\text{-krát}}\Bigg) = \\ &= H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + \sum_{i-1}^{n} \frac{1}{n} H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = \\ &= H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = F(n) + F(m) \end{aligned}$$

■

Veta 2.8. *Nech* $F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$. *Potom* $F(n) = c.\log_2(n)$.

Dôkaz. Indukciou dokážeme $F(n^k) = k.F(n)$ pre $k = 1, 2, \ldots$.
Podľa vety 2.7 platí $F(m.n) = F(m) + F(n)$. Špeciálne pre $m = n$ je $F(n^2) = 2.F(n)$, $F(n^k) = F(n^{k-1}.n) = F(n^{k-1}).F(n) = (k-1).F(n) + F(n) = k.F(n)$.
Môžeme teda písať:

$$F(n^k) = k.F(n) \quad \text{pre } k = 1, 2, \ldots \tag{2.13}$$

Vzťah (2.13) má niekoľko dôsledkov:

1. $F(1) = F(1^2) = 2.F(1)$, čoho vyplýva, že $F(1) = 0$, a teda $F(1) = c.\log_2(1)$ pre každé c.

2. Pretože podľa axiómy AS4 je funkcia F na množine prirodzených čísel rastúca, je pre každé $m > 1 \;\; F(m) > F(1) = 0$.

Vezmime dve prirodzené čísla $m > 1$, $n > 1$ a ľubovoľne veľké prirodzené číslo K. Potom existuje prirodzené číslo k také, že

$$m^k \leq n^K < m^{k+1}. \tag{2.14}$$

Pretože F je rastúca funkcia, je aj

$$F(m^k) \leq F(n^K) < F(m^{k+1}).$$

Použitím (2.13) dostávame

$$k.F(m) \leq K.F(n) < (k+1).F(m).$$

Z posledného výrazu máme ($F(m) > 0$, takže ním možno deliť bez zmeny nerovností)

$$\frac{k}{K} \leq \frac{F(n)}{F(m)} < \frac{k+1}{K}. \tag{2.15}$$

Pretože platí (2.14) môžeme podobnou úvahou postupne písať

$$\log_2(m^k) \leq \log_2(n^K) < \log_2(m^{k+1})$$

$$k.\log_2(m) \leq K.\log_2(n) < (k+1).\log_2(m),$$

a teda (spomeňme si, že $m > 1$ a teda $\log_2(m) > 0$)

$$\frac{k}{K} \leq \frac{\log_2(n)}{\log_2(m)} < \frac{k+1}{K}. \tag{2.16}$$

Ak porovnáme výrazy (2.15) a (2.16) vidíme, že oba zlomky $\frac{F(n)}{F(m)}$, $\frac{\log_2(n)}{\log_2(m)}$ ležia v intervale $\left\langle \frac{k}{K}, \frac{k+1}{K} \right)$ dĺžky $\frac{1}{K}$ a teda

$$\left| \frac{F(n)}{F(m)} - \frac{\log_2(n)}{\log_2(m)} \right| < \frac{1}{K}. \tag{2.17}$$

Celý postup môžeme zopakovať pre ľubovoľne veľké číslo K, a preto (2.17) musí platiť pre ľubovoľné K, čo je možné len tak, že

$$\frac{F(n)}{F(m)} = \frac{\log_2(n)}{\log_2(m)},$$

a teda

$$F(n) = F(m).\frac{\log_2(n)}{\log_2(m)} = \left(\frac{F(m)}{\log_2(m)} \right) \log_2(n). \tag{2.18}$$

Ak v (2.18) fixujeme m a položíme $c = \frac{F(m)}{\log_2(m)}$, dostaneme $F(n) = c.\log_2(n)$. ■

Veta 2.9. *Nech* $p_1 \geq 0,\ p_2 \geq 0, \ldots,\ p_n \geq 0$, $\sum_{i=1}^{n} p_i = 1$. *Potom existuje* $c > 0$ *také, že*

$$H(p_1, p_2, \ldots, p_n) = -c. \sum_{i=1}^{n} p_i . \log_2(p_i). \tag{2.19}$$

Dôkaz. Dokážeme najprv (2.19) pre $p_1, p_2, \ldots, p_n$ racionálne – t. j. v tvare zlomkov dvoch celých nezáporných čísel. Nech s je spoločný menovateľ čísel $p_1, p_2, \ldots, p_n$, nech $p_i = \frac{q_i}{s}$ pre $i = 1, 2, \ldots, n$. Podľa (2.12) vety 2.6 môžeme písať

$$H\Bigg(\underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_1\text{-krát}}, \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_2\text{-krát}}, \ \ldots\ \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_n\text{-krát}}, \Bigg) =$$

$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i . H\left(\frac{1}{q_i}, \frac{1}{q_i} \ldots, \frac{1}{q_i}\right) =$$

$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i . F(q_i) =$$

$$= H(p_1, p_2, \ldots, p_n) + c. \sum_{i=1}^{n} p_i . \log_2(q_i). \quad (2.20)$$

Pretože ľavá strana (2.20) sa rovná $F(s) = c. \log_2(s)$, môžeme písať

$$H(p_1, p_2, \ldots, p_n) = c \log_2(s) - c. \sum_{i=1}^{n} p_i \log_2(q_i) =$$

$$= c \log_2(s) \sum_{i=1}^{n} p_i - c \sum_{i=1}^{n} p_i \log_2(q_i) = c \sum_{i=1}^{n} p_i \log_2(s) - c \sum_{i=1}^{n} p_i \log_2(q_i) =$$

$$= -c \sum_{i=1}^{n} p_i [\log_2(q_i) - \log_2(s)] =$$

$$= -c \sum_{i=1}^{n} p_i \log_2\left(\frac{q_i}{s}\right) = -c \sum_{i=1}^{n} p_i \log_2(p_i). \quad (2.21)$$

Pretože funkcia H je spojitá a (2.21) platí pre všetky racionálne $p_1 \geq 0$, $p_2 \geq 0, \ldots, \ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, musí (2.21) platiť aj pre všetky reálne argumenty p_i spĺňajúce tie isté podmienky. ■

Ostáva nám určiť konštantu c. Aby sme boli v zhode s doterajšími požiadavkami, aby entropia pokusu s dvoma rovnako pravdepodobnými javmi bola jednotková, musí byť $H(1/2, 1/2) = 1$, z čoho vyplýva

$$1 = H\left(\frac{1}{2}, \frac{1}{2}\right) = -c. \left[\frac{1}{2}. \log_2\left(\frac{1}{2}\right) + \frac{1}{2}. \log_2\left(\frac{1}{2}\right)\right] = -c. \left(-\frac{1}{2} - \frac{1}{2}\right) = c$$

Axiomatickou cestou sme sa dostali k tej istej Shannonovej entropickej formule, ako v prípade, keď sme entropiu definovali ako strednú hodnotu diskrétnej náhodnej veličiny informácie.

2.4 Ďalšie vlastnosti entropie

Veta 2.10. *Nech pre všetky* $i = 1, 2, \ldots, n$ *platí* $p_i > 0,\ q_i > 0,\ \sum_{i=1}^{n} p_i = 1,$ $\sum_{i=1}^{n} q_i = 1$. *Potom*

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i), \tag{2.22}$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ *pre všetky* $i = 1, 2, \ldots, n$.

Dôkaz. Najprv dokážeme platnosť nerovnosti

$$\ln(1+y) \leq y \quad \text{pre} \quad y > -1$$

Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy. Je $g'(y) = \dfrac{1}{1+y} - 1$, $g''(y) = -\dfrac{1}{(1+y)^2} \leq 0$. Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$. Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$. Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum. Je preto $g(y) \leq 0$, t. j. $\ln(1+y) - y \leq 0$ a teda $\ln(1+y) \leq y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$. Ak v (2.22) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\ln(x) \leq x - 1 \quad \text{pre} \quad x > 0, \tag{2.23}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.
Dosaďme teraz do (2.23) za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$\begin{aligned}
\ln(q_i) - \ln(p_i) &\leq \frac{q_i}{p_i} - 1 \\
p_i \ln(q_i) - p_i \ln(p_i) &\leq q_i - p_i \\
-p_i \ln(p_i) &\leq -p_i \ln(q_i) + q_i - p_i \\
-\sum_{i=1}^{n} p_i \ln(p_i) &\leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1} \\
-\sum_{i=1}^{n} p_i \frac{\ln(p_i)}{\ln(2)} &\leq -\sum_{i=1}^{n} p_i \frac{\ln(q_i)}{\ln(2)}
\end{aligned}$$

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i),$$

pričom rovnosť v prvých troch riadkoch nastáva práve vtedy, keď $p_i = q_i$, rovnosť v posledných troch riadkoch nastáva práve vtedy, keď $p_i = q_i$ pre všetky $i = 1, 2, \ldots, n$. ■

Veta 2.11. *Pre dané n funkcia*

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

nadobúda maximum pre $p_1 = p_2 = \cdots = p_n = 1/n$.

Dôkaz. Vezmime $p_1, p_2, \ldots, p_n$ ľubovoľné také, že splňujú predpoklady vety, a položme v (2.22) $q_1 = q_2 = \cdots = q_n = \dfrac{1}{n}$. Potom

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2\left(\frac{1}{n}\right) =$$
$$= -\log_2\left(\frac{1}{n}\right) \cdot \sum_{i=1}^{n} p_i = -\log_2(\frac{1}{n}) = \log_2 n = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

■

2.5 Použitie entropie pri riešení niektorých úloh

Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. Predpokladajme, že nastal elementárny jav $\omega \in \Omega$. My nemáme možnosť (a ani potrebu) zistiť, o ktorý elementárny jav ide, nám stačí vedieť, do ktorej množiny rozkladu $\mathbf{B} = \{B_1, B_2, \ldots, B_n\}$ priestoru Ω tento jav padol.[2] Pokus $\mathbf{B} = \{B_1, B_2, \ldots, B_n\}$ na priestore $\Omega, \mathcal{A}, P$ odpovedajúci na otázku, ktorú chceme pokusom zistiť, budeme volať **základný pokus**.

[2] Ak chceme zistiť teplotu snehu kvôli správnemu voskovaniu lyží, stačí zistiť, či je v rozsahoch $(-\infty, -12)$, $(-12, -8)$, $(-8, -4)$, $(-4, 0)$ a $(0, \infty)$, pretože máme k dispozícii vosky určené na vymenované teplotné intervaly.

Často sa vyskytujú úlohy typu: „Zistite na najmenší počet meraní (otázok, skúšok), ktorý z javov základného pokusu **B** nastal.“

Pokiaľ nie je špecifikované inak, predpokladáme, že všetky množiny základného pokusu majú rovnakú pravdepodobnosť. Potom entropia takéhoto pokusu je $\log_2(n)$ – t. j. jeho vykonaním dostaneme informáciu o veľkosti $\log_2(n)$ bitov.

Pre zistenie odpovede na otázku, ktorá nás zaujíma, často nemôžeme zorganizovať priamo základný pokus **B**, napríklad preto lebo počet výsledkov pokusov, ktoré sme schopní urobiť, je obmedzený. Jeden z príkladov obmedzeného počtu možných výsledkov je situácia, keď na otázku môžeme dostať iba dve odpovede – „áno“ alebo „nie“. Ak chceme dostať na našu otázku najväčšiu možnú (strednú) informáciu, musíme ju položiť tak, aby pravdepodobnosť oboch odpovedí bola čo najbližšie k číslu 1/2.

Príklad 2.1. V triede je 32 žiakov, jeden z nich vyhral literárnu súťaž. Ako sa na čo najmenší počet otázok, na ktoré môžeme dostať iba odpovede „áno“ alebo „nie“, zaručene dozvieme, ktorý žiak to bol? Ak by nebolo obmedzenia na počet dovolených odpovedí, problém by plne vyriešil základný pokus s 32 výsledkami, pričom by získaná informácia bola $\log_2(32) = 5$ bitov. Pri dvoch výsledkoch pokusu môžeme získať jednou otázkou najviac 1 bit informácie, takže minimálny počet takýchto otázok na zistenie víťaza je 5.

Ak sme v priemernej slovenskej koedukačnej triede, môžeme položiť otázku: „Je víťaz chlapec?“. Otázka je dobre zvolená, lebo v priemernej slovenskej triede býva približne rovnaký počet chlapcov a dievčat. Odpoveď na takúto otázku prinesie v každom prípade, či odpoveď bude „áno“, alebo „nie“ približne 1 bit informácie.

Otázka „Je víťaz Jano Mrkvička?“ prinesie strednú informáciu veľkosti $H(1/32, 31/32) = -(1/32).\log_2(1/32) - (31/32).\log_2(31/32) = 0.20062$ bitu. Môže sa stať, že odpoveď bude „áno“ a v tom prípade sme získali plných 5 bitov informácie. To sa však stane len v jednom prípade z 32, v ostatných prípadoch dostaneme odpoveď „nie“, a vtedy dostaneme iba 0.0458 bitu informácie. V strednej hodnote prináša otázka typu „Je víťaz Jano Mrkvička?“ 0.2 bitu informácie. Je možné zistiť stredný počet S otázok tohoto typu potrebných na identifikáciu víťaza, avšak ani tento počet otázok S nestačí na zaručené určenie víťaza – na to by bolo treba v najhoršom prípade 31 otázok tohoto typu.

Je preto dobré, aby každá otázka delila žiakov pripadajúcich do úvahy, na dve rovnaké polovice.

Potom je možné dopracovať sa výsledku po piatich otázkach. Postup je nasledujúci: Priradíme žiakom čísla 1 až 32.

1. Otázka „Má víťaz poradové číslo 1 – 16?“ Ak je odpoveď „áno“, vieme že víťaz je v skupine s číslami od 1 do 16, ak je odpoveď „nie“, víťaz je v skupine s číslami od 17 do 32.

2. Otázka „V tej 16-člennej skupine, kde je víťaz, je jeho číslo medzi ôsmimi najmenšími?“ Tým určíme 8-člennú skupinu v ktorej je víťaz.

3. Podobná otázka na štvoricu.

4. Podobná otázka na dvojicu.

5. Otázka na jedného z dvoch.

Predchádzajúci príklad môže vyzerať trochu umelý – ak máme človeka ochotného odpovedať na päť otázok „áno“ alebo „nie“, pravdepodobne bude ochotný odpovedať aj na otázku základného pokusu „Kto vyhral literárnu súťaž?“.

Príklad 2.2. Majme 22 elektrických žiaroviek spojených do série (napr. na vianočný stromček). Jedna žiarovka sa vypálila. Máme po ruke ohmmeter, ktorý môžeme zapojiť do ktoréhokoľvek miesta obvodu. Akým najmenším počtom meraní zaručene nájdeme chybnú žiarovku. Základný pokus má entropiu $\log_2(22) = 4.46$ bitu. Na zistenie zlej žiarovky budeme potrebovať najmenej 5 meraní. Pri prvom meraní zapojíme ohmmeter pred prvú žiarovku a za jedenástu žiarovku. Tým určíme, či je chybná žiarovka medzi prvou a jedenástou, alebo medzi dvanástou až dvadsiatou druhou žiarovkou. Úsek, v ktorom je porucha, rozdelíme na pokiaľ možno rovnaké časti a zase určíme, v ktorej je chybná žiarovka. Po prvom meraní bude podozrivých 11 žiaroviek, po druhom meraní 4 alebo 5 žiaroviek, po treťom meraní 2 alebo 3, po štvrtom meraní už zistíme chybu alebo úsek s dvoma chybnými žiarovkami, a nakoniec po piatom meraní (ak je vôbec potrebné) definitívne určíme chybu.

Príklad 2.3. Máme 27 mincí, z ktorých jedna je falošná - je o máličko ľahšia ako pravá. Máme váhy s dvoma miskami. Na aký najmenší počet vážení možno zaručene určiť falošnú mincu?

Základný pokus má 27 rovnako pravdepodobných výsledkov, jeho entropia je $\log_2(27) = 4.755$ bitov.

Ak dáme na misky váh nerovnaký počet mincí, vždy preváži miska s väčším počtom mincí. Z takéhoto pokusu nedostaneme žiadnu informáciu. Ak dáme

na obe misky rovnaký počet mincí, môžu nastať tri prípady. Aby sme ich mohli ľahšie popísať, označme A množinu mincí na ľavej miske váh, B množinu mincí na pravej miske váh a C množinu ostatných mincí. Ak je falošná minca v množine A, preváži ľavá miska, ak je falošná minca v množine B, preváži pravá miska, ak je falošná minca v množine C, misky budú v rovnováhe. Náš pokus teda odpovie na otázku, v ktorej množine leží falošná minca. Aby sme z pokusu dostali čo najviac informácie, mali by mať množiny A, B, C pokiaľ možno rovnakú pravdepodobnosť. (V našom prípade sa to dá, lebo počet mincí je deliteľný troma). V takom prípade možno jedným vážením dostať informáciu $\log_2(3) = 1.585$ bitu. Keďže $4.755/1.585 = 3$, na vyriešenie úlohy bude treba minimálne tri váženia. Konkrétne riešenie bude nasledovné: Pre prvé váženie rozdelíme mince na tri množiny po 9 mincí. Výsledkom bude identifikácia 9-prvkovej množiny s falošnou mincou. Pri druhom vážení rozdelíme podozrivú množinu 9 mincí na tri podmnožiny po tri mince. Výsledkom druhého pokusu bude identifikácia podozrivej trojprvkovej množiny. Pri treťom vážení dáme na každú misku po jednej minci, jedna minca ostane mimo. Výsledkom posledného váženia bude identifikácia jednej falošnej mince. Stačia nám teda tri váženia.

Práve popísaný postup možno priamo použiť pre hľadanie ľahšej falošnej mince medzi n mincami, ak $n = 3^k$. Ak n nie je deliteľné číslom 3, potom $n = 3m + 1 = m + m + (m + 1)$ – v tom prípade dáme na obe misky váh po m mincí a mimo váh ostane $m + 1$ mincí – t. j. $|A| = |B| = m$, $|C = m + 1|$, alebo $n = 3m + 2 = (m + 1) + (m + 1) + m$ – v tom prípade dáme na obe misky váh po $m + 1$ mincí a mimo váh ostane m mincí – t. j. $|A| = |B| = m + 1$, $|C| = m$.

Príklad 2.4. Majme znovu 27 mincí, z ktorých je jedna falošná – líši sa váhou od pravej. Teraz však nevieme, či je ľahšia alebo ťažšia než pravá minca. Máme určiť nepravú mincu a zistiť, či je ťažšia alebo ľahšia než pravá. Základný pokus má teraz $2 \times 27 = 54$ možných výsledkov – chybná môže byť každá z 27 mincí, pričom môže byť ľahšia alebo ťažšia ako pravá minca. Entropia základného pokusu je $\log_2(54) = 5.755$ bitov, zatiaľ čo entropia jedného pokusu vážením je $\log_2(3) = 1.585$ bitu, z čoho už vidno, že na určenie nepravej mince nemôžu stačiť tri pokusy.

Možný postup riešenia tejto úlohy: Znovu rozdelíme mince na 9 prvkové množiny A, B, C. Nech $w(A)$ je váha množiny A.
1. váženie

a Ak $w(A) = w(B)$, vieme, že falošná minca je v množine C. Druhým vážením zistíme buď $w(A) < w(C)$ – falošná minca je ťažšia, alebo $w(A) > w(C)$ – falošná minca je ľahšia. Tretím vážením určíme trojicu mincí s falošnou mincou a štvrtým vážením samotnú falošnú mincu.

b Ak $w(A) < w(B)$, vieme, množine C sú všetky mince pravé. Druhým vážením zistíme buď $w(A) < w(C)$ – falošná minca je ľahšia a je v množine A, alebo $w(A) > w(C)$ – falošná minca je ťažšia a je v množine A, alebo $w(A) = w(C)$ – falošná minca je ťažšia a je v množine B. Tretím vážením určíme trojicu mincí s falošnou mincou a štvrtým vážením samotnú falošnú mincu.

Príklad 2.5. Majme n veľkých kontajnerov s kovovými guľôčkami. Guľôčky v jednom kontajneri sú rovnako ťažké. $n-1$ kontajnerov obsahuje identické guľôčky, avšak v jednom kontajneri sú guľôčky o 1 gram ťažšie a my nevieme, ktorý kontajner to je. Všetky guľôčky sú zdanlivo rovnaké, takže určiť ťažšiu možno iba vážením. Máme k dispozícii presné dvojmiskové váhy so závažím, ktorými dokážeme odvážiť ľubovoľné množstvo guľôčiek s presnosťou lepšou než 1 gram. Na koľko vážení možno zaručene určiť, v ktorom kontajneri sú ťažšie guľôčky? Základný pokus má n možných výsledkov, jeho entropia je $\log_2(n)$. Ak by sme z každého kontajnera vybrali po guľôčke, dostali by sme problém falošných mincí. Zamyslime sa však, či nemožno pokus zorganizovať tak, aby na jedno váženie dal viac možných výsledkov, ako tri. Pokus urobíme nasledovne: Z prvého kontajera dáme na ľavú misku 1 guľôčku, z druhého 2, atď. až z n-tého kontajnera n guľôčiek. Na pravú misku dáme $1+2+\cdots+n = (1/2)n(n+1)$ guľôčiek z prvého kontajnera. Ak preváži pravá miska, ťažšie guľôčky sú v prvom kontajneri. Ak preváži ľavá miska, dodáme závažie k gramov na vyváženie misiek. Potom ťažšie guľôčky sú v k-tom kontajneri. Na vyriešenie úlohy stačí jedno váženie.

Príklad 2.6. Telefónne vedenie z miesta P do miesta Q je 100 metrov dlhé. Niekde medzi miestami P, Q sa vedenie prerušilo. Vedenie môžeme merať tak, že ho v ľubovoľnom X mieste „napichneme“ a zistíme, či je spojenie medzi miestami P a X. Treba určiť postup s minimálnym počtom meraní, ktorým zaručene identifikujeme úsek vedenia nie dlhší než jeden meter, v ktorom je vedenie prerušené. Ak označíme Y vzdialenosť miesta poruchy X od miesta P, potom Y je spojitá náhodná veličina, $Y \in \langle 0, 100\rangle$. My síce nemáme definovanú entropiu pokusu s nekonečným počtom možných výsledkov, ale intuitívne cítime, že neurčitosť pokusu, ktorý by dával presnú hodnotu Y, je väčšia, než entropia pokusu, ktorý odpovie, v ktorom úseku z n rovnakých úsekov je prerušené vedenie a tá je $H(1/n, 1/n, \ldots, 1/n) = \log_2(n)$. Našťastie, našou úlohou nie je určiť presne miesto chyby, stačí nám úsek s chybou nie dlhší než 1 meter. Ako základný pokus budeme brať pokus

$$\mathbf{B} = \big\{\langle 0,1), \langle 1,2), \ldots, \langle 98,99), \langle 99,100\rangle\big\}$$

s entropiou $H(\mathbf{B}) = \log_2(100) = 6.644$ bitov. Ak už máme identifikovanú chybu v intervale $\langle a, b\rangle$, meranie umožňuje pre ľubovoľné $c \in \langle a, b\rangle$ rozhodnúť, či chyba nastala v intervale $\langle a, c)$ alebo $\langle c, b\rangle$. Ak predpokladáme, že pravdepodobnosť vzniku chyby v nejakom intervale je úmerná jeho dĺžke, na to, aby sme dostali z pokusu čo najväčšiu informáciu, treba voliť polohu bodu c tak, aby delil interval $\langle a, b\rangle$ napoly. Vtedy bude mať takýto pokus entropiu $\log_2(2) = 1$ bit. Pretože pokus $\mathbf{B}$ má entropiu 6.644 bitov, bude treba aspoň 7 meraní. Postup zisťovania chyby bude teda nasledovný: Prvým meraním zistíme, či chyba nastala v prvej alebo druhej polovici vedenia, druhým meraním určíme úsek s chybou dlhý $100/2^2$ m, atď. až šiestym meraním určíme úsek $\langle a, b\rangle$ s chybou dlhý $100/2^6 = 100/64 = 1.5625$ metra. Tento interval obsahuje práve jeden celočíselný bod, do ktorého položíme deliaci bod c pre vykonanie siedmeho posledného merania.

V doterajších úlohách sme hľadali takú organizáciu pokusov, ako pomocou ich minimálneho počtu **zaručene** určiť, ktorý elementárny jav, resp. ktorý jav n-prvkového základného pokusu nastal. Ak je to možné, zorganizujeme ihneď základný pokus. Vo väčšine prípadov problém takýchto úloh spočíva v tom, že sme obmedzení len na pokusy istého typu. Kombináciou minimálneho počtu týchto pokusov máme získať rovnakú informáciu ako základným pokusom. Dolná hranica počtu dovolených pokusov bude priamo úmerná entropii základného pokusu a nepriamo úmerná entropii dovoleného pokusu. Najväčšie rozpaky pred vykonaním pokusov budeme mať vtedy, keď všetky elementárne javy budú mať rovnakú pravdepodobnosť – vtedy neurčitosť pokusu bude $H(1/n, 1/n, \ldots, 1/n) = \log_2(n)$. Toto číslo je maximum informácie, ktoré možno vyťažiť zo základného pokusu a my musíme byť pripravení aj na túto situáciu. Preto predpokladáme rovnakú pravdepodobnosť elementárnych javov základného pokusu. Tento predpoklad navyše vedie k postupu, ktorý nepreferuje žiaden elementárny jav.

Zmenil by sa náš postup v prípade, keby jednotlivé elementárne javy neboli rovnako pravdepodobné? Ak cieľ - nájsť minimálny počet pokusov, ktorým možno zaručene identifikovať elementárny jav, potom nie. Mohli by sme však úlohu preformulovať nasledovne: „Nájsť takú organizáciu pokusov, pri ktorej je stredný počet pokusov minimálny“. Upustili sme od požiadavky „zaručene určiť“. Pripúšťame možnosť, že v niektorých nepriaznivých, ale málo pravdepodobných situáciách bude treba veľa otázok, chceme však aby pri veľkom počte prípadov bol stredný počet pokusov minimálny. Pri takto formulovanej úlohe budeme postupovať inak.

Príklad 2.7. Na štarte automobilových pretekov bolo 18 áut. Z nich autá a_1 a a_2 sú z technologicky vyspelého tímu, a preto pravdepodobnosť víťazstva každého z nich je 1/4. Z ostatných 16 áut $a_3, \ldots, a_{18}$ má každé pravdepodobnosť víťazstva rovnú 1/32. Základný pokus má tvar

$$\mathbf{A} = \{\{a_1\}, \{a_2\}, \{a_2\}, \ldots, \{a_{18}\}\},$$

jeho entropia je

$$H(\mathbf{A}) = H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{32}, \frac{1}{32} \ldots, \frac{1}{32}\right) = 3.5,$$

takže stredný počet otázok s dvomi odpoveďami nemôže byť menší než 3.5. Jedným z možných postupov je v prvej otázke rozhodnúť medzi množinami $A_1 = \{a_1, a_2\}$ a $A_2 = \{a_3, a_4, \ldots, a_{18}\}$. V polovici prípadov vyjde A_1, a tam stačí už len jedna otázka na určenie konkrétneho víťaza. V druhej polovici vyjde A_2 so šestnástimi rovnocennými prvkami, a tu treba na určenie jedného z nich ďalšie 4 otázky. Stredný počet otázok je teda $\frac{1}{2}.2 + \frac{1}{2}.5 = 3.5$.

Nasledujúci postup na zaručené určenie víťaza z 18 prvkovej množiny bude vyžadovať vždy aspoň 4 a v niektorých prípadoch aj 5 otázok

1. Má víťaz číslo medzi 1 – 9? Tým určíme 9-člennú skupinu B_1 obsahujúcu víťaza.
2. Je číslo víťaza medzi štyrmi najnižšími v B_1? Tým určíme štvor- alebo päťčlennú skupinu B_2 s víťazom.
3. Je číslo víťaza medzi dvomi najnižšími v B_2? Tým určíme dvoj- alebo trojčlennú skupinu obsahujúcu víťaza B_3.
4. Je víťaz s najnižším číslom v B_3? Ak áno, STOP, máme víťaza. Inak určíme dvojčlennú skupinu B_4.
5. Priamou otázkou určíme víťaza.

Pojem entropie sa tiež veľmi úspešne využíva pri modelovaní pohybu cestujúcich v skúmanom regióne. Majme v skúmanom regióne n zastávok a chceme určiť pre každú dvojicu zastávok i, j množstvo Q_{ij} ľudí cestujúcich zo zóny i do zastávky j. Priamo možno hodnoty Q_{ij} určiť komplexným dopravným prieskumom, ktorý je však veľmi drahý. Oveľa ľahšie je určiť pre každú zónu i počet P_i cestujúcich, ktorí zo zóny i odchádzajú a tiež počet R_i cestujúcich, ktorí do zóny i prichádzajú. Zrejme platí $\sum_{i=1}^{n} P_i = \sum_{j=1}^{n} R_j = Q$, kde Q je celkový počet cestujúcich za sledované obdobie. Pre neznáme počty Qij musia platiť

nasledujúce rovnosti

$$\sum_{i=1}^{n} Q_{ij} = R_j \quad \text{pre } j = 1, 2, \ldots, n \tag{2.24}$$

$$\sum_{j=1}^{n} Q_{ij} = P_i \quad \text{pre } i = 1, 2, \ldots, n \tag{2.25}$$

$$Q_{ij} \geq 0 \quad \text{pre } i,\ j = 1, 2, \ldots, n \tag{2.26}$$

Nech c_{ij} sú náklady na prepravu jedného cestujúceho zo zóny i do zóny j. (Tieto náklady obsahujú nielen cestovné, ale aj časové straty cestujúceho, nepohodlie cestovania atď.) Jednou z hypotéz je, že dopravný systém sa ustáli tak, že sa budú minimalizovať celkové dopravné náklady

$$C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} Q_{ij}. \tag{2.27}$$

Za predpokladu tejto hypotézy možno hodnoty Q_{ij} dostať minimalizáciou (2.27) za predpokladov (2.24), (2.25) a (2.26), čo je problém známy ako klasická vybilancovaná dopravná úloha. Žiaľ, práve popísaný model dáva výsledky značne odlišné od praktických pozorovaní.

Ukazuje sa však, že v rámci rovnakej spoločensko - ekonomickej situácie je rovnaká miera slobody voľby východiska a cieľa, ktorá sa dá vyjadriť entropiou

$$H\left(\frac{Q_{11}}{Q}, \frac{Q_{12}}{Q}, \ldots \frac{Q_{1n}}{Q}, \frac{Q_{21}}{Q}, \frac{Q_{22}}{Q}, \ldots \frac{Q_{2n}}{Q}, \ldots\ldots, \frac{Q_{n1}}{Q}, \frac{Q_{n2}}{Q}, \ldots \frac{Q_{nn}}{Q}\right). \tag{2.28}$$

Podiel $\dfrac{Q_{ij}}{Q}$ v (2.28) totiž vyjadruje pravdepodobnosť, že cestujúci cestuje zo zastávky i do zastávky j.

Entropické modely založené na maximalizácii (2.28) resp. na kombinácii kriteriálnych funkcií (2.27) a (2.28), alebo rozšírením ohraničení o jedno z $C \leq C_0$ alebo $H \geq H_0$ už lepšie zodpovedajú praktickým skúsenostiam.

2.6 Podmienená entropia

Označme $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ nejaký pokus na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$. Predpokladajme, že nastal elementárny jav $\omega \in \Omega$. Chceme vedieť, ktorý z javov B_j nastal, t. j. pre ktoré $j = 1, 2, \ldots m$ je $\omega \in B_j$. Pre nejaké ohraničenia nemôžeme vykonať pokus $\mathbf{B}$ (tým skôr sa nemôžeme dozvedieť, ktorý jav $\omega \in \Omega$ nastal), ale dozvieme sa výsledok pokusu $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$. Predpokladajme, že jeho výsledkom je jav A_i. Ak už vieme, že nastal jav A_i, javy $B_1,\ B_2,\ \ldots,\ B_m$ nastanú s pravdepodobnosťami $P(B_1|A_i)$, $P(B_2|A_i)$, $\ldots$, $P(B_m|A_i)$. Neurčitosť pokusu $\mathbf{B}$ sa zmení z hodnoty

$$H(\mathbf{B}) = H(P(B_1), P(B_2), \ldots, P(B_m)$$

na hodnotu

$$H\big(P(B_1|A_i), P(B_2|A_i), \ldots, P(B_m|A_i)\big),$$

ktorú budeme označovať $H(\mathbf{B}|A_i)$.

Definícia 2.3. Nech $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ sú dva pokusy. **Podmienenou entropiu pokusu B za predpokladu, že nastal jav** A_i (alebo len za podmienky A_i) je

$$H(\mathbf{B}|A_i) = H\big(P(B_1|A_i), P(B_2|A_i), \ldots, P(B_m|A_i)\big) = \\ = -\sum_{j=1}^{m} P(B_j|A_i).\log_2(P(B_j|A_i)). \quad (2.29)$$

Príklad 2.8. Hádžeme hracou kockou. Označme $\mathbf{B} = \{B_1, B_2, \ldots, B_6\}$ pokus, v ktorom jav B_i znamená „padlo i bodov“ pre $i = 1, 2, \ldots 6$. Všetky javy majú rovnakú pravdepodobnosť $P(B_i) = 1/6$.
Naša neurčitosť o výsledku pokusu $\mathbf{B}$ je

$$H(\mathbf{B}) = H(1/6, 1/6, \ldots, 1/6) = \log_2(6) = 2.585 \text{ bitu.}$$

Predpokladajme, že sa po uskutočnení pokusu dozvieme, že padlo nepárne číslo. Označme $A_1 = B_1 \cup B_3 \cup B_5$, $A_2 = B_2 \cup B_4 \cup B_6$. Jav A_1 znamená „padlo nepárne číslo“, jav A_2 znamená „padlo párne číslo“. Správa „padlo nepárne číslo“ nesie so sebou $-\log_2(P(A_1)) = -\log_2(1/2) = 1$ bit informácie.

Po správe „padlo nepárne číslo“ sa naša neurčitosť o výsledku pokusu zmení z $H(\mathbf{B})$ na

$$H(\mathbf{B}|A_1) = \\ H\big(P(B_1|A_1), P(B_2|A_1), P(B_3|A_1), P(B_4|A_1), P(B_5|A_1), P(B_6|A_1)\big) = \\ H(1/3, 0, 1/3, 0, 1/3, 0) = H(1/3, 1/3, 1/3) = \log_2(3) = 1.585 \text{ bitu.}$$

Správa „padlo nepárne číslo“ – t. j. jav A_1 nesúca 1 bit informácie znžila našu neurčitosť o výsledku pokusu z $H(\mathbf{B}) = 2.585$ na $H(\mathbf{B}|A_1) = 1.585$ – práve o množstvo informácie, ktoré so sebou niesla. POZOR! Toto nie je všeobecne platná skutočnosť. Nasledujúci príklad ukáže, že v niektorých prípadoch správa „nastal jav A_i“ o výsledku pokusu $\mathbf{A}$ môže priniesť dokonca zvýšenie podmienenej entropie $H(\mathbf{B}|A_i)$.

Príklad 2.9. Michail Schumacher bol fenomenálny pilot formuly 1, ktorý získal v rokoch 1994, 1995 a 2000–2004 sedem titulov majstra sveta. V roku 2004 vyhral 13 pretekov z celkového počtu 18, takže pravdepodobnosť jeho víťazstva bola takmer 3/4. Na základe tejto skutočnosti vytvorme nasledujúci modelový príklad.

Na štarte je 17 jazdcov – Schumacher s pravdepodobnosťou víťazstva 3/4 a ďalších 16 rovnocenných jazdcov, z ktorých má každý šancu na víťazstvo rovnajúcu sa 1/64. Označme $\mathbf{B} = \{B_1, B_2, \ldots, B_{17}\}$ pokus, v ktorom B_1 je jav znamenajúci, že vyhral Schumacher, B_i pre $i = 2, 3, \ldots, 17$ je jav, že vyhral i-ty jazdec. Nech $P(B_1) = 3/4$, $P(B_2) = P(B_3) = \cdots = P(B_{17}) = 1/64$.

Ak sa po skončení preteku dozvieme, vyhral Schumacher, dostaneme $-\log_2(P(B_1)) = -\log_2(0.75) = 0.415$ bitov informácie. Ak sa však dozvieme, že vyhral jazdec číslo 17, dostaneme $-\log_2(P(B_{17})) = -\log_2(1/64) = 6$ bitov informácie. Entropia pokusu $\mathbf{B}$ je

$$H(\mathbf{B}) = H\left(3/4, 1/64, 1/64, \ldots, 1/64\right) = 1.811.$$

Majme pokus $\mathbf{A} = \{A_1, A_2\}$, kde A_1 je jav „vyhral Schumacher“ a A_2 je jav „nevyhral Schumacher“. Je $P(A_1) = 3/4$, $P(A_2) = 1/4$. Predpokladajme, že sa po preteku dozvieme, že tentokrát Schumacher nevyhral – nastal jav A_2. Táto správa nesie so sebou $-\log_2(P(A_2)) = -\log_2(1/4) = 2$ bity informácie. Naša neurčitosť po tejto správe sa zmení a $H(\mathbf{B}) = 1.811$ na $H(\mathbf{B}|A_2)$. Počítajme

$$H(\mathbf{B}|A_2) = H\big(P(B_1|A_2), P(B_2|A_2), \ldots, P(B_{17}|A_2)\big) = \\ = H(0, 1/16, 1/16, \ldots, 1/16) = H(1/16, 1/16, \ldots, 1/16) = 4.$$

Správa „nastal jav A_2“ (t. j. „Schumacher nevyhral“) doniesla 2 bity informácie, a napriek tejto správe naša neurčitosť o výsledku preteku stúpla z $H(\mathbf{B}) = 1.811$ na hodnotu $H(\mathbf{B}|A_2) = 4$.

Výsledok A_2 pokusu **A** nastane s pravdepodobnosťou 1/4, výsledok A_1 má pravdepodobnosť 3/4 a vtedy $H(\mathbf{B}|A_1) = 0$. Stredná hodnota zvyškovej neurčitosti pokusu **B** po vykonaní pokusu **A** sa bude rovnať

$$P(A_1).H(\mathbf{B}|A_1) + P(A_2).H(\mathbf{B}|A_2) = (3/4).0 + (1/4).4 = 1.$$

Stredná hodnota zvyškovej entropie pokusu **B** po vykonaní pokusu **A** bude 1 bit.

Zrekapitulujme si v krátkosti úvahu pred definíciou 2.3. Zaujímame sa o výsledok pokusu **B**, ktorý má entropiu $H(\mathbf{B})$. Nastal elementárny jav $\omega \in \Omega$; my sme však dostali správu, že $\omega \in A_i$ a táto správa zmenila entropiu pokusu **B** na $H(\mathbf{B}|A_i)$. Ku každému $\omega \in \Omega$ existuje práve jedna množina $A_i \in \mathbf{A}$ taká, že $\omega \in A_i$. Môžeme teda každému $\omega \in \Omega$ priradiť jednoznačne číslo $H(\mathbf{B}|A_i)$ – toto priradenie je diskrétnou náhodnou veličinou[3] na Ω. Stredná hodnota tejto náhodnej veličiny je $\sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i)$.

Definícia 2.4. Nech sú dané dva pokusy

$$\mathbf{A} = \{A_1, A_2, \ldots, A_n\}, \quad \mathbf{B} = \{B_1, B_2, \ldots, B_m\}.$$

Podmienenou entropiu pokusu B za predpokladu, vykonania pokusu A (alebo len za podmienky **A**) je

$$H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i). \tag{2.30}$$

[3]Presne by sme mohli definovať túto náhodnú veličinu ako

$$h(\mathbf{B}|\mathbf{A})(\omega) = \sum_{i=1}^{n} H(\mathbf{B}|A_i).\chi_{A_i}(\omega),$$

kde $\chi_{A_i}(\omega)$ je indikátor množiny A_i, t. j. $\chi_{A_i}(\omega) = 1$ práve vtedy, keď $\omega \in A_i$, inak $\chi_{A_i}(\omega) = 0$.

Platí:

$$\begin{aligned}\sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i) &= \sum_{i=1}^{n} P(A_i).H\big(P(B_1|A_i), P(B_2|A_i), \ldots, P(B_m|A_i)\big) = \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j|A_i).\log_2(P(B_j|A_i)) = \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).\frac{P(A_i\cap B_j)}{P(A_i)}.\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right) = \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right).\end{aligned}$$

Môžeme teda tiež písať

$$H(\mathbf{B}|\mathbf{A}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right) \tag{2.31}$$

Čím bude hodnota $H(\mathbf{B}|A_i)$ menšia, tým presnejšie jav A_i charakterizuje výsledok pokusu $\mathbf{B}$. Ak teda navrhujeme pokus $\mathbf{A}$ ako jeden z pokusov, ktorých minimálny počet zaručene dospieť k určeniu javu B_i, potom sa treba snažiť navrhnúť pokus $\mathbf{A}$ tak, aby maximum z hodnôt $H(\mathbf{B}|A_i)$ bolo minimálne.

Definícia 2.5. Nech $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ sú dva pokusy na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$. Potom **kombinovaným pokusom pokusov A, B** nazveme pokus

$$\mathbf{A}\wedge\mathbf{B} = \big\{A_i\cap B_j \mid A_i\in\mathbf{A},\ B_j\in\mathbf{B}\big\} \tag{2.32}$$

Ak najprv vykonáme pokus $\mathbf{A}$ a potom pokus $\mathbf{B}$, (alebo aj najprv $\mathbf{B}$ a potom $\mathbf{A}$), dozvieme sa to isté, t. j. získame rovnakú informáciu, ako keby sme vykonali pokus $\mathbf{A}\wedge\mathbf{B}$. Ak už vykonáme pokus $\mathbf{A}$ a jeho výsledok je A_i, podmienená entropia pokusu $\mathbf{B}$ za predpokladu, že nastal jav A_i, je $H(\mathbf{B}|A_i)$. Keďže jav A_i má pravdepodobnosť $P(A_i)$, jeho príspevok k celkovej strednej hodnote pokusu $\mathbf{B}$ za predpokladu, že je známy výsledok pokusu $\mathbf{A}$, je $P(A_i).H(\mathbf{B}|A_i)$ a podmienená entropia pokusu $\mathbf{B}$ za predpokladu, že poznáme výsledok pokusu $\mathbf{A}$, je $H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i)$.

Podľa vety 2.6 (str. 28) platí vzťah (2.12). Vezmime pokus $\mathbf{A} \wedge \mathbf{B}$. Označme $q_{ij} = P(A_i \cap B_j)$, $p_i = P(A_i)$. Potom platí

$$p_i = P(A_i) = \sum_{j=1}^{m} p(A_i \cap B_j) = \sum_{j=1}^{m} q_{ij}.$$

Predpoklady vety 2.6 sú teda splnené a preto je

$$\begin{aligned} H(\mathbf{A} \wedge \mathbf{B}) &= H\left(\underbrace{q_{11}, q_{12}, \dots q_{1m}}_{p_1}, \underbrace{q_{21}, q_{22}, \dots q_{2m}}_{p_2}, \dots, \underbrace{q_{n1}, q_{n2}, \dots q_{nm}}_{p_n}\right) = \\ &= H(p_1, p_2, \dots, p_n) + \sum_{j=1}^{m} p_i . H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \dots, \frac{q_{im}}{p_i}\right) = \\ &= H\big(P(A_1), P(A_2), \dots, P(A_n)\big) + \\ &+ \sum_{i=1}^{m} P(A_i) H\left(\frac{P(A_i \cap B_1)}{P(A_i)}, \frac{P(A_i \cap B_2)}{P(A_i)}, \dots, \frac{P(A_i \cap B_m)}{P(A_i)}\right) = \\ &= H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \end{aligned}$$

Teda platí nasledujúca veta:

Veta 2.12.

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \tag{2.33}$$

Podľa vzťahu (2.33) je $H(\mathbf{B}|\mathbf{A})$ zvyšková entropia kombinovaného pokusu $\mathbf{A} \wedge \mathbf{B}$ po vykonaní pokusu $\mathbf{A}$. Vidíme tiež, že o čo je entropia $H(\mathbf{A})$ pokusu $\mathbf{A}$ väčšia, o to menšia je podmienená entropia $H(\mathbf{B}|\mathbf{A})$.

Definícia 2.6. Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ sú dva pokusy na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$. Hovoríme, že pokusy $\mathbf{A}$, $\mathbf{B}$ sú **štatisticky nezávislé** (alebo len nezávislé), ak pre každé $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ sú A_i, B_j nezávislé javy.

2.7 Spoločná informácia pokusov

Znovu sa vráťme k situácii, keď sa zaujímame sa o výsledok pokusu **B** s entropiou $H(\mathbf{B})$. Tento pokus však z nejakých dôvodov nemôžeme vykonať, ale vykonáme pokus **A**. V situácii, keď už poznáme výsledok pokusu **A**, neurčitosť pokusu **B** sa zmení z $H(\mathbf{B})$ na $H(\mathbf{B}|\mathbf{A})$ – toto je stredné množstvo dodatočnej informácie, ktorú možno získať z pokusu **B** po vykonaní pokusu **A**. Rozdiel $H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$ možno považovať za stredné množstvo informácie o pokuse **B** obsiahnuté v pokuse **A**.

Definícia 2.7. Stredné množstvo informácie $I(\mathbf{A}, \mathbf{B})$ o pokuse B v pokuse A je

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \tag{2.34}$$

Veta 2.13.

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \tag{2.35}$$

Dôkaz. Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A} \wedge \mathbf{B}) - H(\mathbf{A})$ z (2.33) do (2.34) dostaneme žiadaný vzťah. ■

Zo vzťahu (2.35) vidíme, že $I(\mathbf{A}, \mathbf{B}) = I(\mathbf{B}, \mathbf{A})$, t. j., že informácia o pokuse **B** obsiahnutá v pokuse **A** sa rovná informácii o pokuse **A** obsiahnutej v pokuse **B**. Preto sa niekedy hodnote $I(\mathbf{A}, \mathbf{B})$ hovorí aj **spoločná informácia pokusov A, B**.

Veta 2.14. *Nech* $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$I(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i \cap B_j) . \log_2 \left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)} \right) . \tag{2.36}$$

Dôkaz. Pretože $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$ je rozklad priestoru Ω je

$$B_j = B_j \cap \Omega = B_j \cap \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i \cap B_j .$$

Pretože zjednotenie na pravej strane posledného výrazu je disjunktné, je

$$P(B_j) = \sum_{i=1}^{n} P(A_i \cap B_j) .$$

Dosadením za $H(\mathbf{B}|\mathbf{A})$ zo vzťahu (2.31) do definičnej rovnosti $I(\mathbf{A},\mathbf{B})$ dostávame

$$
\begin{aligned}
&I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) = \\
&= -\sum_{j=1}^{m} P(B_j).\log_2 P(B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right) = \\
&= -\sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i\cap B_j).\log_2 P(B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right) = \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\left[\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i)}\right) - \log_2 P(B_j)\right] = \\
&\qquad = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i\cap B_j)}{P(A_i).P(B_j)}\right)
\end{aligned}
$$

■

Veta 2.15. *Nech* $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$0 \le I(\mathbf{A},\mathbf{B}), \tag{2.37}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ *štatisticky nezávislé.*

Dôkaz. Použijeme vzťah (2.36) z vety 2.14 a nerovnosť $\ln x \le x-1$, ktorá platí pre všetky $x > 0$, pričom rovnosť nastáva práve vtedy, keď $x = 1$.

$$
\begin{aligned}
&P(A_i\cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i\cap B_j)}\right) = P(A_i\cap B_j).\ln(2).\ln\left(\frac{P(A_i).P(B_j)}{P(A_i\cap B_j)}\right) \le \\
&\le P(A_i\cap B_j).\ln(2).\left[\left(\frac{P(A_i).P(B_j)}{P(A_i\cap B_j)}\right) - 1\right] = \ln(2).\left[P(A_i).P(B_j) - P(A_i\cap B_j)\right],
\end{aligned}
$$

pričom rovnosť platí práve vtedy, keď $\dfrac{P(A_i).P(B_j)}{P(A_i\cap B_j)} = 1$, t. j. vtedy, keď sú javy A_i, B_j nezávislé.

Použitím práve dokázanej nerovnosti máme

$$\begin{aligned}
-I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) \leq \\
&\leq \ln(2).\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P(A_i).P(B_j) - P(A_i \cap B_j)\right)\right] = \\
&= \ln(2).\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j)}_{=1}\right] = \\
&= \ln(2).\left[\sum_{i=1}^{n} P(A_i)\underbrace{\sum_{j=1}^{m} P(B_j)}_{=1} - 1\right] = \ln(2).\left[\underbrace{\sum_{i=1}^{n} P(A_i)}_{=1} - 1\right] = 0,
\end{aligned}$$

pričom rovnosť platí práve vtedy, keď sú všetky dvojice javov A_i, B_j pre $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ nezávislé. ■

Veta 2.16.

$$H(\mathbf{B}|\mathbf{A}) \leq H(\mathbf{B}), \tag{2.38}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ *štatisticky nezávislé.*

Dôkaz. Tvrdenie vety vyplýva zo vzťahu $0 \leq I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$, v ktorom rovnosť nastáva, len keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ štatisticky nezávislé. ■

Veta 2.17.

$$H(\mathbf{A} \wedge \mathbf{B}) \leq H(\mathbf{A}) + H(\mathbf{B}), \tag{2.39}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ *štatisticky nezávislé.*

Dôkaz. Podľa (2.35) vety 2.13 a podľa vety 2.15 je $0 \leq I(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$, pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ štatisticky nezávislé. ■

Kapitola 3

Zdroje informácie

3.1 Reálne zdroje informácie

Objekt (osoba, zariadenie, prístroj), ktorý je schopný na svojom výstupe produkovať nejaký signál, budeme volať zdroj informácie. Zdrojom informácie môže byť človek signalizujúci baterkou znaky Morseovej abecedy, klávesnica počítača, vysielajúca 8–bitové slová, telefónny prístroj produkujúci analógový signál od 300 do 3400 hertzov, signál z prehrávača kompaktných diskov produkujúci 44000 16-bitových vzoriek za sekundu, televízny obrazový signál obsahujúci 25 obrázkov za sekundu atď.

Vidíme, že televízny obrazový signál bude neporovnateľne zložitejší, než telefónny signál. Ale každý uzná, že desať minút sledovania obrazovky s monoskopom, prenášaného týmto zložitým signálom nedá toľko informácie, koľko desať minút telefonického rozhovoru.

Zdroje informácie môžu produkovať spojitý alebo diskrétny signál. Každý spojitý signál však možno v diskrétnych časových okamihoch odmerať – diskrétne vzorkovať, a pokiaľ je vzorkovacia frekvencia dvojnásobná ako maximálna frekvencia signálu, stačia tieto diskrétne vzorky na plnohodnotnú rekonštrukciu pôvodného signálu.

Môžeme teda predpokladať, že zdroj produkuje v časových okamihoch $t = t_1, t_2, t_3, \ldots$ signály $X_{t_1}, X_{t_2}, X_{t_3}, \ldots$, ktoré môžeme považovať za diskrétne náhodné veličiny – nadobúdajú len konečne veľa rôznych hodnôt. Konečnú množinu rôznych diskrétnych signálov produkovaných zdrojom nazveme abecedou zdroja, jednotlivé prvky abecedy zdroja nazveme znakmi.

Časové okamihy $t = t_1, t_2, t_3, \ldots$ môžu, ale nemusia byť pravidelné. Napríklad zdroj vysielajúci v Morseovej abecede používa symboly bodka, čiarka, krátka medzera (oddeľuje bodky a čiarky v rámci jedného písmena), dlhá medzera (oddeľuje od seba jednotlivé písmená). V inej interpretácii môžeme oddeľovaciu medzeru medzi bodkami a čiarkami v rámci jedného písmena považovať za súčasť bodky a čiarky, a v tomto prípade máme zdroj produkujúci tri znaky a to bodky, čiarky a medzery. V oboch interpretáciách však časové okamihy, $t = t_1, t_2, t_3, \ldots$, v ktorých sa vysielajú jednotlivé symboly, nie sú rovnaké – vyslanie bodky trvá kratšie, než vyslanie čiarky.

Je výhodné považovať časový interval medzi dvoma za sebou nasledujúcimi časovými okamihmi za jednotkový – potom pracujeme s náhodnými veličinami $X_1, X_2, X_3 \ldots$.

Definícia 3.1. Diskrétny náhodný proces je postupnosť náhodných veličín $\mathcal{X} = X_1, X_2, X_3 \ldots$. Ak X_i nadobudne hodnotu a_i pre $i = 1, 2, \ldots$, postupnosť $a_1, a_2, \ldots$ nazveme **realizáciou náhodného procesu** $\mathcal{X}$.

V tejto kapitole budeme skúmať informačnú výdatnosť rôznych zdrojov informácie. Zdroje informácie sa od seba líšia frekvenciou, s akou sú schopné vysielať, počtom znakov abecedy zdroja – hodnôt, ktoré môžu nadobúdať náhodné veličiny X_i – a tiež ich pravdepodobnostným rozdelením. Aby sme sa zbavili vplyvu frekvencie zdroja, budeme sa snažiť charakterizovať zdroje podľa množstva informácie pripadajúce na jeden vyslaný znak. Avšak ani frekvencia výstupu znakov zo zdroja, ani mohutnosť výstupnej abecedy zdroja neurčuje plne množstvo informácie, ktoré produkuje zdroj. To bude značne závisieť aj od rozdelenia pravdepodobnosti náhodných veličín X_i.

3.2 Matematický model informačného zdroja

Definícia 3.2. Nech X je konečná množina, nech X^* je množina všetkých konečných postupností prvkov z X vrátane prázdnej postupnosti, ktorú budeme značiť symbolom e. Množinu X nazveme **abecedou**, jej prvky **znakmi abecedy** X, prvky množiny X^* nazveme **slovami**, e **prázdnym slovom**. Označme X^n množinu všetkých n-prvkových postupností znakov z X, jej prvky nazveme **slovami dĺžky** n. Nech $P : X^* \to \mathbb{R}$ je reálna nezáporná funkcia definovaná na X^* s nasledujúcimi vlastnosťami:

1. $P(e) = 1$ (3.1)
2. $\sum_{(x_1,\ldots,x_n)\in X^n} P(x_1,\ldots,x_n) = 1$ (3.2)
3. $\sum_{(y_{n+1},\ldots,y_{n+m})\in X^m} P(x_1,\ldots,x_n,y_{n+1},\ldots,y_{n+m}) = P(x_1,\ldots,x_n)$ (3.3)

Potom usporiadanú dvojicu $\mathcal{Z} = (X^*, P)$ nazveme **zdrojom informácie** alebo krátko **zdrojom**. Číslo $P(x_1, x_2, \ldots, x_n)$ nazveme **pravdepodobnosťou slova** $x_1, \ldots, x_n$.

Číslo $P(x_1, x_2, \ldots, x_n)$ vyjadruje pravdepodobnosť toho, že zdroj od svojho okamihu spustenia vyšle v čase 1 znak x_1, v čase 2 znak x_2, atď., až v čase n vyšle znak x_n, čo sa dá povedať i tak, že $P(x_1, x_2, \ldots, x_n)$ je pravdepodobnosť vyslania slova $x_1, x_2, \ldots, x_n$ za n časových okamihov od spustenia zdroja. Podmienka (3.1) hovorí, že za 0 časových okamihov vyšle zdroj prázdne slovo s pravdepodobnosťou rovnajúcou sa 1, druhá podmienka hovorí, že za n časových okamihov od spustenia zdroj s istotou vyšle nejaké slovo z abecedy X. Tretia podmienka sa volá podmienkou konzistencie a vyjadruje požiadavku, aby pravdepodobnosť množiny všetkých slov dĺžky $n+m$ začínajúcich slovom $x_1, x_2, \ldots, x_n$ sa rovnala pravdepodobnosti $P(x_1, x_2, \ldots, x_n)$, lebo

$$\{y_1, y_2, \ldots, y_{n+m} \mid y_1 = x_1, y_2 = x_2, \ldots, y_n = x_n\} = \\ = \bigcup_{z_1,z_2\ldots,z_m\in X^m} \{x_1, x_2, \ldots, x_n, z_1, z_2, \ldots, z_m\} \ .$$

Na tomto mieste je potrebné pripomenúť dva rozdiely medzi chápaním pojmu „slovo“ v lingvistickom a našom zmysle. V lingvistickom zmysle slovo napr. slovenského jazyka je taká postupnosť písmen, ktorá je prijatá do množiny slov – slovníka slovenského jazyka. Tak napríklad slovo „víkend“ je slovom slovenského jazyka na rozdiel od slova „weekend“. V našom informatickom zmysle sú chápané ako slová všetky konečné postupnosti znakov abecedy X, „víkend“, „weekend“, „dnekeew“, „qwdíýážťfj“ – to všetko sú slová abecedy $X = \{a,á,b,c,č,\ldots,z,ž\}$. Druhým podstatným rozdielom je, že slová v prirodzenom jazyku sa oddeľujú medzerou, na rozdiel od našej definície, v ktorej je výstup zo zdroja možné chápať ako jedno (možno aj veľmi dlhé) slovo, ale súčasne aj niekoľko bezprostredne po sebe nasledujúcich slov, ktoré nie sú od seba ničím oddelené, resp. výstupné slovo si môžeme podeliť na slová v ľubovoľných miestach ako nám je to pre naše účely výhodné.

Zaujíma nás pravdepodobnosť $P_n(y_1, y_2 \ldots, y_m)$, s akou zdroj vyšle slovo $y_1, y_2 \ldots, y_m$ v čase n, presnejšie v časových okamihoch $n, n+1, \ldots, n+m-1$. Túto pravdepodobnosť vypočítame nasledovne:

$$P_n(y_1, y_2, \ldots, y_m) = \sum_{(x_1,\ldots,x_{n-1})\in X^{n-1}} P(x_1, x_2, \ldots, x_{n-1}, y_1, y_2, \ldots, y_m) \ . \quad (3.4)$$

Definícia 3.3. Hovoríme, že zdroj $\mathcal{Z} = (X^*, P)$ je **stacionárny**, ak pravdepodobnosti $P_i(x_1, x_2, \ldots, x_n)$ nezávisia od i, t. j. ak

$$P_i(x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_n) \quad \text{pre každé } i \text{ a každé } x_1, x_2 \ldots, x_n \in X^n \ .$$

Označme X_i diskrétnu náhodnú premennú, ktorá bude popisovať vyslanie jedného znaku zo zdroja $\mathcal{Z} = (X^*, P)$ v čase i. Potom jav „v čase i zdroj vyslal znak x“ je vlastne javom $[X_i = x]$ a teda $P([X_i = x]) = P_i(x)$. Vyslanie slova $x_1, x_2, \ldots, x_n$ v čase i možno pomocou náhodných veličín X_i modelovať ako jav $[X_i = x_1] \cap [X_{i+1} = x_2] \cap \cdots \cap [X_{i+n-1} = x_n]$, čo skrátene zapíšeme $[X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n]$, z čoho máme

$$P([X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n]) = P_i(x_1, x_2, \ldots, x_n).$$

Definícia 3.4. Hovoríme, že zdroj $\mathcal{Z} = (X^*, P)$ je **nezávislý**, ak pre ľubovoľné $i,\ j,\ n,\ m$ také, že $i + n \le j$ platí

$$P\Big([X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n] \cap \\ \cap [X_j = y_1, X_{j+1} = y_2, \ldots, X_{j+m-1} = y_m]\Big) = \\ = P\big([X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n]\big). \\ .P\big([X_j = y_1, X_{j+1} = y_2, \ldots, X_{j+m-1} = y_m]\big).$$

Zdroj je nezávislý, ak vyslanie ľubovoľného slova v ľubovoľnom čase j nezávisí od toho, čo zdroj vyslal do času j. Niekedy sa takýmto zdrojom hovorí aj bezpamäťové.

Zdroj vysielajúci stať v slovenskom jazyku nie je nezávislý zdroj. Ako uvádza Černý v [5] je veľa slovenských slov, obsahujúcich „ZA“, ale žiadne slovo neobsahuje podslovo „ZAZA“. Je teda $P(ZA) > 0$, a v prípade nezávislosti by malo byť $P(ZAZA) = P(ZA).P(ZA) > 0$, ale $P(ZAZA) = 0$. Gramatické pravidlo o zhode predmetu a prívlastku spôsobí, že podmienená pravdepodobnosť výskytu slova „␣CHLAPA“ za predpokladu, že ho bezprostredne predchádza

slovo „ÉHO“ (medzeru „␣“ považujeme tiež za symbol abecedy) bude oveľa väčšia než podmienená pravdepodobnosť výskytu toho istého slova za predpokladu, že ho bezprostredne predchádza slovo „ÉMU“.

Z krátkodobého hľadiska by sme s istým priblížením mohli považovať písanú slovenčinu za stacionárny zdroj, avšak z hľadiska storočí badať aj tu zmeny – čoraz menej sa používajú niektoré slová ako rínok, kantár, pitvor, merica, dieža a začínajú sa používať nové slová ako víkend, mobil, procesor, internet. Stacionarita zdroja je však jedným zo základných predpokladov, za ktorých môžeme dostať v teórii informácie použiteľné výsledky, preto odteraz budeme predpokladať, že zdroje, s ktorými pracujeme, sú stacionárne. Tento predpoklad je v praktických situáciách splnený aj v prípade prirodzených jazykov, pretože my ich používame v krátkych časových intervaloch.

3.3 Entropia zdroja

Majme stacionárny zdroj $\mathcal{Z} = (Z^*, P)$. Nech má abeceda zdroja m symbolov, t. j. nech $Z = \{a_1, a_2, \ldots, a_m\}$. Chceme vedieť, akú strednú informáciu dostaneme, keď sa dozvieme, aký znak zdroj vyslal. Vyslanie znaku v ľubovoľnom čase možno pri stacionárnom zdroji považovať za vykonanie pokusu

$$\mathbf{B} = \big\{\{a_1\}, \{a_2\}, \ldots, \{a_m\}\big\}$$

s pravdepodobnosťami $p_1 = P(a_1),\ p_2 = P(a_2),\ \ldots,\ p_m = P(a_m)$. Entropia tohoto pokusu je $H(\mathbf{B}) = H(p_1, p_2, \ldots, p_m)$, čo je stredná hodnota informácie získanej týmto pokusom.

Skúmajme teraz informáciu, ktorú dostaneme v dvoch po sebe idúcich znakoch vyslaných zo stacionárneho zdroja $\mathcal{Z} = (Z^*, P)$. Príslušný pokus bude teraz

$$\mathbf{C}_2 = \big\{(a_{i_1}, a_{i_2}) \mid a_{i_1} \in Z,\ a_{i_2} \in Z\big\}\ .$$

Pokus $\mathbf{B}$ môžeme prezentovať aj ako

$$\mathbf{B} = \big\{\{a_1\} \times Z,\ \{a_2\} \times Z,\ \ldots,\ \{a_m\} \times Z\big\}\ .$$

Ak definujeme $\mathbf{D} = \big\{Z \times \{a_1\},\ Z \times \{a_2\},\ \ldots,\ Z \times \{a_m\}\big\}$, potom $\mathbf{C}_2 = \mathbf{B} \wedge \mathbf{D}$ a $H(\mathbf{D}) = H(\mathbf{B}) = H(p_1, p_2, \ldots, p_m)$.
Podľa vety 2.17 (str. 48) platí

$$H(\mathbf{C}_2) = H(\mathbf{B} \wedge \mathbf{D}) \leq H(\mathbf{B}) + H(\mathbf{D}) = 2.H(\mathbf{B})\ .$$

Teraz túto vlastnosť rozšírime na n-znakové slová. Budeme postupovať matematickou indukciou.
Predpokladajme, že pre pokus

$$\mathbf{C}_n = \left\{(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \ldots, n\right\}$$

už platí $H(\mathbf{C}_n) \leq n.H(\mathbf{B})$. Pokus $\mathbf{C}_n$ má rovnakú entropiu ako pokus

$$\mathbf{C}'_n = \left\{(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) \times Z \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \ldots, n\right\} .$$

Označme

$$\mathbf{C}_{n+1} = \left\{(a_{i_1}, a_{i_2}, \ldots, a_{i_{n+1}}) \mid a_{i_k} \in Z, \text{ pre } k = 1, 2, \ldots, n+1\right\},$$
$$\mathbf{D} = \left\{Z^n \times \{a_1\},\ Z^n \times \{a_2\},\ \ldots,\ Z^n \times \{a_m\}\right\},$$

potom

$$H(\mathbf{C}_{n+1}) = H(\mathbf{C}'_n \wedge \mathbf{D}) \leq H(\mathbf{C}'_n) + H(\mathbf{D}) \leq n.H(\mathbf{B}) + H(\mathbf{B}) = (n+1).H(\mathbf{B}) .$$

Tým sme dokázali, že pre všetky prirodzené n platí $H(\mathbf{C}_n) < n.H(\mathbf{B})$.

Vidíme, že v prípade stacionárneho zdroja, ktorý nie je nezávislý, je priemerná entropia na jedno písmeno $\frac{1}{n}H(\mathbf{C}_n)$ vždy menšia ako entropia prvého písmena. To nás vedie k myšlienke, definovať entropiu zdroja ako priemernú entropiu na jedno písmeno pre veľmi dlhé slová.

Definícia 3.5. Nech $\mathcal{Z} = (Z^*, P)$ je zdroj informácie. Nech existuje limita

$$H(\mathcal{Z}) = -\lim_{n\to\infty} \frac{1}{n} . \sum_{(x_1,\ldots,x_n)\in Z} P(x_1, x_2, \ldots, x_n).\log_2(P(x_1, x_2, \ldots, x_n)). \tag{3.5}$$

Potom číslo $H(\mathcal{Z})$ nazveme **entropiou zdroja** $\mathcal{Z}$.

Pre stacionárny nezávislý zdroj $\mathcal{Z} = (Z^*, P)$ vypočítame entropiu ľahko. Platí totiž nasledujúca veta.

Veta 3.1. *Nech (Z^*, P) je stacionárny nezávislý zdroj. Potom*

$$H(\mathcal{Z}) = -\sum_{x\in Z} P(x).\log_2 P(x). \tag{3.6}$$

Dôkaz. Platí:

$$\sum_{(x_1,\ldots,x_n)\in Z} P(x_1,x_2,\ldots,x_n).\log_2(P(x_1,x_2,\ldots,x_n)) =$$

$$= \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\big[\log_2 P(x_1)+\log_2 P(x_2)+\cdots+\log_2 P(x_n)\big] =$$

$$= \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_1)+$$

$$+ \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_2)+$$

$$+ \ldots\ldots\ldots\ldots\ldots\cdots+$$

$$+ \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_n) =$$

$$= \sum_{x_1\in Z} P(x_1).\log_2 P(x_1)\ .\ \underbrace{\sum_{(x_2,\ldots,x_n)\in Z} P(x_2).P(x_3),\ldots,P(x_n)}_{=1} + \cdots =$$

$$= \sum_{x_1\in Z} P(x_1).\log_2 P(x_1) + \sum_{x_2\in Z} P(x_2).\log_2 P(x_2) + \cdots + \sum_{x_3\in Z} P(x_3).\log_2 P(x_3) =$$

$$= n.\sum_{x\in Z} P(x).\log_2 P(x).$$

Z posledného výrazu už vyplýva tvrdenie vety. ■

Ak je zdroj len stacionárny, limita (3.5) vôbec nemusí existovať.

Veta 3.2. Shannon – Mac Millan. *Nech $\mathcal{Z} = (Z^*, P)$ je stacionárny nezávislý zdroj. Potom k ľubovoľnému $\varepsilon > 0$ existuje prirodzené číslo $n(\varepsilon)$ také, že pre všetky $n \geq n(\varepsilon)$ je*

$$P\left\{x_1,\ldots x_n \in Z^n \mid \left|\frac{1}{n}.\log_2 P(x_1,\ldots x_n) + H(\mathcal{Z})\right| \geq \varepsilon\right\} < \varepsilon\ . \tag{3.7}$$

Túto vetu uvádzame v jej najjednoduchšej forme a bez dôkazu. Platí však aj pre oveľa všeobecnejšie zdroje (napríklad ergodické zdroje), ku ktorým možno s istým priblížením zaradiť aj slovenčinu i ostatné prirodzené jazyky. Spomenuté všeobecnejšie zdroje je však bez použitia aparátu teórie miery (pozri prístup uvedený vo voliteľnej časti 3.5) ťažko presnejšie špecifikovať a študovať. Všeobecnejšie formulácie Shannonovej – Mac Millanovej vety s použitím najjednoduchšieho matematického aparátu nájde čitateľ v knihe [9].
Označme

$$E(n,\varepsilon) = \left\{ x_1, \dots x_n \in Z^n \mid \left| \frac{1}{n}.\log_2 P(x_1, \dots x_n) + H(\mathcal{Z}) \right| < \varepsilon \right\} \tag{3.8}$$

Shannonova – Mac Millanova veta hovorí, že pre každé $\varepsilon > 0$ existuje množina $E(n,\varepsilon)$, pre ktorú platí $P(E(n,\varepsilon)) > 1-\varepsilon$.
Platí:

$$\begin{aligned}(x_1,\dots,x_n) \in E(n,\varepsilon) &\iff -\varepsilon < \frac{1}{n}\log_2 P(x_1,\dots,x_n) + H(\mathcal{Z}) < \varepsilon \iff \\ &\iff -n(H(\mathcal{Z})+\varepsilon) < \log_2 P(x_1,\dots,x_n) < -n(H(\mathcal{Z})-\varepsilon) \iff \\ &\iff 2^{-n(H(\mathcal{Z})+\varepsilon)} < P(x_1,\dots,x_n) < 2^{-n(H(\mathcal{Z})-\varepsilon)}\end{aligned}$$

Nech $|E(n,\varepsilon)|$ je počet prvkov v množine $E(n,\varepsilon)$. Pretože pravdepodobnosť každého prvku množiny $E(n,\varepsilon)$ je väčšia než $2^{-n(H(\mathcal{Z})+\varepsilon)}$, je

$$1 \geq P(E(n,\varepsilon)) > |E(n,\varepsilon)|.2^{-n(H(\mathcal{Z})+\varepsilon)}.$$

Na druhej strane je pravdepodobnosť každého prvku množiny $E(n,\varepsilon)$ menšia než $2^{-n(H(\mathcal{Z})-\varepsilon)}$, z čoho

$$1-\varepsilon < P(E(n,\varepsilon)) < |E(n,\varepsilon)|.2^{-n(H(\mathcal{Z})-\varepsilon)}.$$

Z týchto nerovností dostávame nasledujúce ohraničenia:

$$(1-\varepsilon).2^{n(H(\mathcal{Z})-\varepsilon)} < |E(n,\varepsilon)| < 2^{n(H(\mathcal{Z})+\varepsilon)} \tag{3.9}$$

Množina všetkých slov dĺžky n sa teda rozpadne na významnú množinu $E(n,\varepsilon)$, ktorá má približne $2^{n.H(\mathcal{Z})}$ slov, ktorých pravdepodobnosť sa málo líši od $2^{H(\mathcal{Z})}$, a na zvyšok slov s bezvýznamnou celkovou pravdepodobnosťou.

Slovenčina používa 26 písmen abecedy bez diakritiky a 15 písmen s diakritikou á, č, ď, é, í, ľ, Í, ň, ó, ô, š, ť, ú, ý, ž. Navyše sa používajú aj interpunkčné znamienka (čiarka, dvojbodka, bodkočiarka, pomlčka, úvodzovky, bodka, výkričník, otáznik a medzera). Aj keď prijmeme námietku, že slovenčina by vystačila bez písmen q, w, x, potrebuje jej abeceda Z minimálne 40 znakov. (A to ešte nepoužívame veľké písmená.) Entropia slovenčiny určite neprevýši číslo 2. Počet všetkých 8-znakových slov abecedy Z je teda 40^8, $|E(8,\varepsilon)|$ odhadneme na $2^{8.2} = 2^{16}$.

$$\text{Je } \frac{2^{16}}{40^8} = 6.10^{-8}$$

Množina $E(8,\varepsilon)$ významných 8-znakových slov obsahuje približne 6 milióntin percenta počtu všetkých 8-znakových slov.

3.4 Produkt informačných zdrojov

Definícia 3.6. Majme dva informačné zdroje $\mathcal{Z}_1 = (A^*, P_1)$, $\mathcal{Z}_2 = (B^*, P_2)$. **Produktom zdrojov** $\mathcal{Z}_1$**,** $\mathcal{Z}_2$ nazveme zdroj $\mathcal{Z}_1 \times \mathcal{Z}_2 = ((A \times B)^*, P)$, kde $(A \times B)$ je karteziánskym súčinom množín A a B (t. j. množinou všetkých usporiadaných dvojíc (a, b), kde $a \in A$, $b \in B$) a kde $P(e) = 1$ (pravdepodobnosť vyslania prázdneho slova za 0 časových okamihov) a kde

$$P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) = P(a_1,a_2,\ldots,a_n).P(b_1,b_2,\ldots,b_n) \quad (3.10)$$

pre ľubovoľné $a_i \in A$, $b_j \in B$, $i, j \in \{1, 2, \ldots, n\}$.

Veta 3.3. *Produkt zdrojov* $\mathcal{Z}_1$*,* $\mathcal{Z}_2$ *je korektne definovaný, t. j. pre pravdepodobnosť* P *platí* (3.1)*,* (3.2)*,* (3.3) *z definície zdroja 3.2.*

Dôkaz. Vzťahy (3.1), (3.2), (3.3) z definície zdroja 3.2 (str. 50) prepíšeme nasledovne:

1. $P(e) = 1$ (3.11)
2. $$\sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) = 1 \quad (3.12)$$
3. $$\sum_{(p_1,q_1),\ldots,(p_m,q_m)\in(A\times B)^m} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n),(p_1,q_1),\ldots,(p_m,q_m)\big) = $$
$$= P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) \quad (3.13)$$

Prvý vzťah vyplýva z definície 3.2 zdroja $\mathcal{Z}_1 \times \mathcal{Z}_2$. Dokážeme platnosť tretieho vzťahu.

$$\sum_{(p_1,q_1),\ldots,(p_m,q_m)\in(A\times B)^m} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n),(p_1,q_1),\ldots,(p_m,q_m)\big) =$$
$$= \sum_{p_1p_2\ldots p_m\in A^m} \sum_{q_1q_2\ldots q_m\in B^m} P(a_1,\ldots,a_n,p_1,\ldots,p_m).P(b_1,\ldots,b_n,q_1,\ldots,p_m) =$$
$$= \sum_{p_1p_2\ldots p_m\in A^m} P(a_1,\ldots,a_n,p_1,\ldots,p_m) \sum_{q_1q_2\ldots q_m\in B^m} P(b_1,\ldots,b_n,q_1,\ldots,p_m) =$$
$$= P(a_1,a_2,\ldots,a_n)\ .\ P(b_1,b_2,\ldots,b_n) = P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big)\ .$$

Analogicky sa dokáže jednoduchší druhý vzťah. ■

Veta 3.4. *Majme dva informačné zdroje $\mathcal{Z}_1$, $\mathcal{Z}_2$ s entropiami $H(\mathcal{Z}_1)$, $H(\mathcal{Z}_2)$. Potom pre entropiu zdroja $\mathcal{Z}_1 \times \mathcal{Z}_2$ platí*

$$H(\mathcal{Z}_1 \times \mathcal{Z}_2) = H(\mathcal{Z}_1) + H(\mathcal{Z}_2)\ . \tag{3.14}$$

Dôkaz.

$$H(\mathcal{Z}_1 \times \mathcal{Z}_2) =$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P\big((a_1,b_1),\ldots,(a_n,b_n)\big).\log_2 P\big((a_1,b_1),\ldots,(a_n,b_n)\big) =$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} \Big\{ P(a_1,\ldots,a_n).P(b_1,\ldots,b_n).$$
$$. \left[\log_2 P(a_1,\ldots,a_n) + \log_2 P(b_1,\ldots,b_n)\right] \Big\} =$$
$$= \lim_{n\to\infty} \frac{1}{n} \left[\sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P(a_1,\ldots,a_n).P(b_1,\ldots,b_n).\log_2 P(a_1,\ldots,a_n) + \right.$$
$$\left. + \sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P(a_1,\ldots,a_n).P(b_1,\ldots,b_n).\log_2 P(b_1,\ldots,b_n) \right] =$$

$$= \lim_{n\to\infty} \frac{1}{n} \left[\sum_{a_1,\ldots,a_n \in A^n} P(a_1,\ldots,a_n).\log_2 P(a_1,\ldots,a_n) \; . \underbrace{\sum_{b_1,\ldots,b_n \in B^n} P(b_1,\ldots,b_n)}_{=1} + \right.$$

$$\left. + \sum_{b_1,\ldots,b_n \in B^n} P(b_1,\ldots,b_n) \log_2 P(b_1,\ldots,b_n) \; . \underbrace{\sum_{a_1,\ldots,a_n \in A^n} P(a_1,\ldots,a_n)}_{=1} \right] =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{a_1,\ldots,a_n \in A^n} P(a_1,\ldots,a_n).\log_2 P(a_1,\ldots,a_n) +$$

$$+ \lim_{n\to\infty} \frac{1}{n} \sum_{b_1,\ldots,b_n \in B^n} P(b_1,\ldots,b_n) \log_2 P(b_1,\ldots,b_n) = H(\mathcal{Z}_1) + H(\mathcal{Z}_2).$$

■

Definícia 3.7. Nech $\mathcal{Z} = (A^*, P)$ je informačný zdroj. Definujme $\mathcal{Z}^2 = \mathcal{Z} \times \mathcal{Z}$ a ďalej indukciou $\mathcal{Z}^n = \mathcal{Z}^{n-1} \times \mathcal{Z}$.

Zdroj $\mathcal{Z}^n = \underbrace{\mathcal{Z} \times \mathcal{Z} \times \cdots \times \mathcal{Z}}_{n\text{-krát}}$ je zdroj s abecedou A^n. Použitím vety 3.4 a matematickej indukcie dostaneme nasledujúcu vetu:

Veta 3.5. *Nech $\mathcal{Z}$ je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}^n)$ zdroja $\mathcal{Z}^n$ platí*

$$H(\mathcal{Z}^n) = n.H(\mathcal{Z}) \tag{3.15}$$

Definícia 3.8. Nech $\mathcal{Z} = (A^*, P)$. Označme $\mathcal{Z}_{(k)} = \left((A^k)^*, P_{(k)}\right)$ zdroj s abecedou A^k, kde $P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$ pre $\mathbf{a}_i \in A^k$, $\mathbf{a}_i = a_{i1}a_{i2}\ldots a_{ik}$ je definované ako

$$P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = P(a_{11}, a_{12}, \ldots, a_{1k}, a_{21}, a_{22}, \ldots, a_{2k}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nk})$$

Informačný zdroj $\mathcal{Z}_{(k)}$ vznikne z informačného zdroja $\mathcal{Z}$ tak, že zo zdroja $\mathcal{Z}$ budeme odoberať každý k-ty okamih celé výstupné slovo dĺžky k v pôvodnej abecede, pričom budeme výstupné k-znakové slová brať ako znaky novej abecedy.

Pozor! Je podstatný rozdiel medzi $\mathcal{Z}_{(k)}$ a $\mathcal{Z}^k$. Kým výstupné slová zdroja $\mathcal{Z}_{(k)}$ sú k-tice po sebe idúcich znakov pôvodného zdroja $\mathcal{Z}$ a ich znaky môžu byť medzi sebou závislé, slová zdroja $\mathcal{Z}^k$ vznikli ako k-tice výstupov k navzájom nezávislých identických zdrojov so zdrojom $\mathcal{Z}$ a znaky v rámci jednotlivých výstupných slov sú navzájom nezávislé.
V prípade stacionárneho nezávislého zdroja $\mathcal{Z}$ je však $\mathcal{Z}_{(k)} \equiv \mathcal{Z}^k$.

Veta 3.6. *Nech $\mathcal{Z}$ je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}_{(k)})$ zdroja $\mathcal{Z}_{(k)}$ platí*

$$H(\mathcal{Z}_{(k)}) = k.H(\mathcal{Z}) \tag{3.16}$$

Dôkaz. Platí

$$\begin{aligned} H(\mathcal{Z}_{(k)}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{\mathbf{a}_1,\ldots,\mathbf{a}_n \in A^n} P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{a_{ij}\in A \text{ pre } 1\le i\le n,\ 1\le j\le k} P(a_{11}, \ldots, a_{1k}, a_{21}, \ldots, a_{2k}, \ldots, a_{n1}, \ldots, a_{nk}) = \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{x_1,x_2,\ldots,x_{n.k}\in A} P(x_1, x_2, \ldots, x_{n.k}) = \\ &= k. \left[\lim_{n\to\infty} \frac{1}{k.n} \sum_{x_1,x_2,\ldots,x_{n.k}\in A} P(x_1, x_2, \ldots, x_{n.k}) \right] = k.H(\mathcal{Z}) \end{aligned} \tag{3.17}$$

■

Posledná veta hovorí, že stredná informácia na jedno k-znakové slovo – znak zdroja $\mathcal{Z}_{(k)}$ – je k-násobkom strednej informácie pripadajúcej na jeden znak pôvodného zdroja $\mathcal{Z}$. Je to očakávaná skutočnosť – „uzátvorkovaním“ výstupu zdroja $\mathcal{Z}$ po k znakov sa stredná informácia pripadajúca na jeden znak pôvodnej abecedy zdroja $\mathcal{Z}$ nezmení.

3.5 Informačný zdroj ako súčin priestorov s mierou*

Aj keď pomocou modelu z predchádzajúcej kapitoly dokážeme definovať a odvodiť mnohé potrebné vlastnosti zdrojov, má tento model isté nedostatky. Jedným z nich je, že systém funkcií $P(x_1, x_2, \ldots, x_n)$ sa nedá interpretovať ako pravdepodobnostná miera na množine Z^* všetkých slov abecedy Z.

Existuje model, ktorý tento nedostatok nemá, vyžaduje však použitie aparátu teórie miery. Táto časť je založená na teórii rozširovnia miery a teórii súčinu priestorov s mierou – 3. a 7. kapitola knihy [6] a na výsledkoch ergodickej teórie – [3].
Nech $Z = \{a_1, a_2, \ldots, a_r\}$. Označme

$$\Omega = \prod_{i=-\infty}^{\infty} Z \tag{3.18}$$

množinu všetkých postupností prvkov z množiny Z tvaru

$$\boldsymbol{\omega} = (\ldots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \ldots) \ .$$

Na množine Ω definujeme pre každé celé číslo i funkciu X_i predpisom

$$X_i(\boldsymbol{\omega}) = \omega_i \ .$$

Nech $E_1, E_2, \ldots E_k$ sú podmnožiny množiny Z. **Cylindrom** nazveme množinu

$$C_n(E_1, E_2, \ldots, E_k) = \\ = \{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) \in E_1,\ X_{n+1}(\boldsymbol{\omega}) \in E_2, \ldots,\ X_{n+k-1}(\boldsymbol{\omega}) \in E_k\}.$$

Nech $x_1, x_2, \ldots, x_k$ je ľubovoľná konečná postupnosť prvkov z množiny Z. **Elementárnym cylindrom** nazveme množinu

$$EC_n(x_1, x_2, \ldots, x_k) = \\ = \{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) = x_1,\ X_{n+1}(\boldsymbol{\omega}) = x_2, \ldots,\ X_{n+k-1}(\boldsymbol{\omega}) = x_k\}.$$

Všimnime si, že možno písať

$$C_n(E_1, E_2, \ldots, E_k) = \cdots \times Z \times Z \times E_1 \times E_2 \times \cdots \times E_k \times Z \times Z \times \ldots,$$

resp.

$$EC_n(x_1, x_2, \ldots, x_k) = \cdots \times Z \times Z \times \{x_1\} \times \{x_2\} \times \cdots \times \{x_k\} \times Z \times Z \times \ldots$$

Elementárny cylinder $EC_n(x_1, x_2, \ldots, x_k)$ predstavuje situáciu, kedy zdroj v čase n až $n+k-1$ vyšle slovo $(x_1, x_2, \ldots, x_k)$.

Označme $\mathcal{F}_0$ množinu všetkých cylindrov. Množina $\mathcal{F}_0$ obsahuje prázdnu množinu (napríklad cylinder $C_1(\emptyset)$ je prázdny), obsahuje Ω (pretože $C_1(Z) = \Omega$), je uzavretá na konečné prieniky a na konečne zjednotenia. Preto existuje najmenšia σ-algebra $\mathcal{F}$ podmnožín priestoru Ω obsahujúca $\mathcal{F}_0$. Pozri [6], (kapitola 7 – Product Spaces).

Definícia 3.9. Informačným zdrojom s abecedou Z nazveme pravdepodobnostný priestor $\mathcal{Z} = (\Omega, \mathcal{F}, P)$, kde $\Omega = \prod_{i=-\infty}^{\infty} Z$, $\mathcal{F}$ je najmenšia σ-algebra podmnožín priestoru Ω obsahujúca všetky cylindre a kde P je nejaká pravdepodobnostná miera na σ-algebre $\mathcal{F}$.

Pretože každý cylinder možno napísať ako zjednotenie elementárnych cylindrov, stačilo by definovať $\mathcal{F}$ ako najmenšiu σ-algebru obsahujúcu všetky elementárne cylindre.

Definíciou 3.9 sme dosiahli, čo sme chceli. Máme pravdepodobnostný priestor, v ktorom je vyslanie ľubovoľného slova v ľubovoľnom čase javom (elementárnym cylindrom) a na ktorom sa dajú všeobecne modelovať a študovať rôzne vlastnosti zdrojov.

Definícia 3.10. Nech $(\Omega, \mathcal{F}, P)$ je pravdepodobnostný priestor, nech $T : \Omega \to \Omega$ je vzájomne jednoznačné zobrazenie na Ω. Pre $A \subseteq \Omega$ označme

$$T^{-1}A = \{\omega \mid T(\omega) \in A\} \qquad T(A) = \{T(\omega) |\ \omega \in A\}. \tag{3.19}$$

Indukciou môžeme definovať $T^{-n}A$ takto: $T^{-1}A$ je definované v (3.19). Majme definované $T^{-n}A$, potom definujeme $T^{-(n+1)}A = T^{-1}(T^{-n}A)$.
Hovoríme, že zobrazenie T je **merateľné**, ak pre každé $A \in \mathcal{F}$ je $T^{-1}A \in \mathcal{F}$.
Hovoríme, že zobrazenie T **zachováva mieru**, ak T je merateľné zobrazenie a ak pre každé $A \in \mathcal{F}$ je $P(T^{-1}A) = P(A)$.
Hovoríme, že T je **premiešavajúce zobrazenie**, ak T je vzájomne jednoznačné, zachováva mieru a pre ľubovoľné dve množiny $A, B \in \mathcal{F}$ je

$$\lim_{n\to\infty} P(A \cap\ T^{-n}B) = P(A).P(B). \tag{3.20}$$

Hovoríme, že množina $B \in \mathcal{F}$ je T**-invariantná**, množina ak $T^{-1}B = B$.
Hovoríme, že T je **ergodické zobrazenie**, ak T je vzájomne jednoznačné, zachováva mieru a všetky jeho invariantné množiny majú mieru 0 alebo 1.

Veta 3.7. *Nech T je premiešavajúce zobrazenie. Potom T je ergodické zobrazenie.*

Dôkaz. T je vzájomne jednoznačné a zachováva mieru. Treba dokázať, že jediné T-invariantné množiny sú množiny miery 0 alebo 1. Nech $B \in \mathcal{F}$ je T -invariantná, nech $A \in \mathcal{F}$ je ľubovoľná merateľná množina. Potom

$$\begin{aligned}
\lim_{n\to\infty} P(A \cap\ T^{-1}B) &= P(A).P(B) \\
P(A \cap B) &= P(A).P(B) \quad \text{pre každé } A \in \mathcal{F} \\
P(B \cap B) &= P(B).P(B) \\
P(B) &= (P(B))^2 \\
(P(B))^2 - P(B) &= 0 \\
P(B)[1 - P(B)] &= 0
\end{aligned}$$

Z poslednej rovnosti už vyplýva že buď $P(B) = 0$, alebo $P(B) = 1$.

Veta 3.8. Ergodická veta. *Nech T je ergodické zobrazenie na pravdepodobnostnom priestore $(\Omega, \mathcal{F}, P)$. Potom pre každú množinu $A \in \mathcal{F}$ a pre skoro všetky*[1] *$\omega \in \Omega$ platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_A\left(T^i(\omega)\right) = P(A), \tag{3.21}$$

kde $\chi_A(\omega)$ je indikátor množiny A, t. j. $\chi_A(\omega) = 1$, ak $\omega \in A$, inak $\chi_A(\omega) = 0$.

Dôkaz. Dôkaz vety je zložitý, čitateľ ho nájde v [3]. ■

Definícia 3.10 a vety 3.7, 3.8 platia pre všeobecné pravdepodobnostné priestory. Vráťme sa teraz k informačnému zdroju $\mathcal{Z} = (\Omega, \mathcal{F}, P)$, kde Ω je množinou všetkých z oboch strán nekonečných postupností znakov konečnej abecedy Z. Na množine Ω definujeme vzájomne jednoznačné zobrazenie T, ktoré nazveme posun predpisom

$$X_n(T(\boldsymbol{\omega})) = X_{n+1}(\boldsymbol{\omega}) \tag{3.22}$$

$$\begin{aligned}
\boldsymbol{\omega} &= \ldots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \ldots \\
T(\boldsymbol{\omega}) &= \ldots, \omega_{-1}, \omega_0\ \ , \omega_1, \omega_2, \omega_3, \ldots
\end{aligned}$$

Zobrazenie T „posunie“ postupnosť znakov $\boldsymbol{\omega}$ o jedno miesto doľava.

[1] Termín „pre skoro všetky $\omega \in \Omega$“ znamená: pre všetky $\omega \in \Omega - \phi$, kde $\phi \subset \Omega$ má nulovú pravdepodobnostnú mieru.

Je ešte jeden pohľad na zobrazenie T. Nech $T^n(\boldsymbol{\omega})$ je n-krát aplikované zobrazenie T, teda

$$T^n(\boldsymbol{\omega}) = \underbrace{T(T(\ldots T(\boldsymbol{\omega})\ldots))}_{n\text{-krát}}.$$

Exaktne možno definovať indukciou $T^1(\boldsymbol{\omega}) = T(\boldsymbol{\omega})$, $T^{n+1}(\boldsymbol{\omega}) = T(T^n(\boldsymbol{\omega}))$. $X_0(\boldsymbol{\omega})$ je znak postupnosti $\boldsymbol{\omega}$ vyslaný zdrojom v čase 0, $X_0(T(\boldsymbol{\omega}))$ je znak postupnosti $\boldsymbol{\omega}$ vyslaný zdrojom v čase 1, $X_0(T^2(\boldsymbol{\omega}))$ je znak postupnosti $\boldsymbol{\omega}$ vyslaný zdrojom v čase 2 atď.
Majme cylinder $C_n(E_1, E_2, \ldots, E_k)$, potom

$$T^{-1}C_n(E_1, E_2, \ldots, E_k) = C_{n+1}(E_1, E_2, \ldots, E_k),$$

$$T^{-m}C_n(E_1, E_2, \ldots, E_k) = C_{n+m}(E_1, E_2, \ldots, E_k).$$

Vlastnosti zobrazenia T spolu s pravdepodobnostnou mierou P podstatne charakterizujú vlastnosti zdroja, preto za zdroj môžeme považovať štvoricu $\mathcal{Z} = (\Omega, \mathcal{F}, P, T)$.

Definícia 3.11. Hovoríme, že zdroj $\mathcal{Z} = (\Omega, \mathcal{F}, P, T)$ je **stacionárny**, ak posun T je mieru zachovávajúce zobrazenie.

Veta 3.9. *Nech $\mathcal{F}_0$ je algebra generujúca σ-algebru $\mathcal{F}$. Nech $T^{-1}A \in \mathcal{F}_0$ a $P(T^{-1}A) = P(A)$ pre každé $A \in \mathcal{F}_0$. Potom je T mieru zachovávajúce zobrazenie.*

Dôkaz predchádzajúcej vety vyžaduje znalosť postupov teórie miery, preto ho neuvádzame. Čitateľ ho môže nájsť v knihe [3]. Pre prístup k modelovaniu zdrojov pomocou aparátu teórie miery je však táto veta typická tým, že v mnohých prípadoch stačí dokázať nejakú vlastnosť miery alebo zobrazenia T pre prvky generujúcej algebry $\mathcal{F}_0$ a prostriedky teórie miery dokážu tieto vlastnosti pre všetky javy zo σ-algebry, generovanej algebrou $\mathcal{F}_0$. Dôsledkom tejto vety je, že pre dôkaz stacionarity zdroja $\mathcal{Z}$ stačí ukázať, že posun T zachováva mieru cylindrov.

Príklad 3.1. Majme zdroj $\mathcal{Z} = (\Omega, \mathcal{F}, P, T)$ s konečnou abecedou

$$Z = \{a_1, a_2, \ldots, a_r\}.$$

Nech sú dané pravdepodobnosti $p_1 = P(a_1)$, $p_2 = P(a_2), \ldots$, $p_r = P(a_r)$, $\sum_{i=1}^{r} p_i = 1$. Mieru P definujeme množinou jej hodnôt na všetkých elementárnych cylindrov predpisom

$$P\big(EC_n(a_{i_1}, a_{i_2}, \ldots, a_{i_k})\big) = p_{i_1}.p_{i_2}, \ldots, p_{i_k}. \tag{3.23}$$

Túto mieru ľahko rozšírime na algebru $\mathcal{F}_0$ všetkých cylindrov

$$P\big(C_n(E_1, E_2, \ldots, E_k)\big) = P(E_1).P(E_2).\ldots.P(E_k). \tag{3.24}$$

Veta 3.10. *Nech $\mathcal{F}_0$ je algebra generujúca $\mathcal{F}$. Ak $T^{-1}A \in \mathcal{F}$ a $P(T^{-1}A) = P(A)$ pre všetky $A \in \mathcal{F}_0$, potom je T zobrazenie zachovávajúce mieru.*
Teória miery teda zaručuje existenciu jedinej miery P na $\mathcal{F}$ spĺňajúcej (3.23). Niekedy sa zobrazenie T na práve popísanom pravdepodobnostnom priestore nazýva Bernouliho posun. Príslušný zdroj je stacionárnym a nezávislým zdrojom. Otázkou je, či je Bernouliho posun ergodickým zobrazením. Vezmime dva cylindre $A = C_s(E_1, E_2, \ldots, E_k)$, $B = C_t(F_1, F_2, \ldots, F_l)$. Ak je n dostatočne veľké, bude mať množina $A \cap T^{-n}A$ tvar

$$\begin{aligned} &A \cap T^{-n}B = \\ &= \cdots \times Z \times Z \times E_1 \times E_2 \times \cdots \times E_k \times Z \times \cdots \times Z \times F_1 \times F_2 \times \cdots \times F_l \times Z \times Z \ldots \end{aligned}$$

čo je vlastne cylinder $C_s(E_1, E_2, \ldots, E_k, Z, \ldots, Z, F_1, F_2, \ldots, F_l)$, ktorého pravdepodobnosť je podľa (3.24) $\prod_{i=1}^{k} P(E_i).\prod_{j=1}^{l} P(F_j) = P(A).P(B)$. Ak A, B sú cylindre, máme

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A).P(B) \tag{3.25}$$

Opäť si pomôžeme vetou z teórie miery.

Veta 3.11. *Nech $\mathcal{F}_0$ je algebra generujúca σ-algebru $\mathcal{F}$. Ak* (3.25) *platí pre všetky $A, B \in \mathcal{F}_0$, potom je zobrazenie T premiešavajúce.*

Bernouliho posun je premiešavajúce a teda aj ergodické zobrazenie.

Príklad 3.2. Nech Ω, $\mathcal{F}$, T sú ako v predchádzajúcom príklade, ale nech P je teraz všeobecná pravdepodobnostná miera na $\mathcal{F}$ taká že sa zachováva pri zobrazení T. Povedať, že T zachováva mieru P je ekvivalentné s tvrdením, že

$$P\{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) \in E_1, X_{n+1}(\boldsymbol{\omega}) \in E_2, \ldots, X_{n+k-1}(\boldsymbol{\omega}) \in E_k, \} \tag{3.26}$$

nezávisí na n, čo je ekvivalentné s definíciou stacionarity náhodného procesu $\{X_i\}_{i=-\infty}^{\infty}$.
Pretože množina všetkých konečných zjednotení disjunktných elementárnych cylindrov tvorí algebru generujúcu σ-algebru $\mathcal{F}$, je miera P na $\mathcal{F}$ jednoznačne určená svojimi hodnotami

$$P(x_1, x_2, \ldots, x_k) = \\ = P\{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) = x_1, X_{n+1}(\boldsymbol{\omega}) = x_2, \ldots, X_{n+k-1}(\boldsymbol{\omega}) = x_k\} \quad (3.27)$$

(Príslušné tvrdenia o rozšírení miery nájde čitateľ v [6], kap. 3 a kap. 7, a v [3].) Pretože musí platiť (3.26) a pretože T zachováva mieru, musí byť

$$1. \quad P(x_1, x_2, \ldots, x_k) \geq 0 \quad (3.28)$$

$$2. \quad \sum_x P(x_1, x_2, \ldots, x_k, x) = P(x_1, x_2, \ldots, x_k) \quad (3.29)$$

$$3. \quad \sum_x P(x) = 1 \quad (3.30)$$

$$4. \quad \sum_x P(x, x_1, x_2, \ldots, x_k) = P(x_1, x_2, \ldots, x_k) \quad (3.31)$$

Naopak, ak máme systém funkcií $P(x_1, \ldots, x_k)$ spĺňajúcich (3.28), (3.29), (3.30) a (3.31), potom existuje jediná miera P na $\mathcal{F}$, ktorá sa zachováva pri zobrazení T a pre ktorú platí (3.27).

Definícia 3.12. Hovoríme, že matica $\mathbf{\Pi} = (q_{ij})$ typu $r \times r$ je **stochastická matica**, ak $q_{ij} \geq 0$ pre všetky i, j také, že $1 \leq i \leq r$, $1 \leq j \leq r$ a platí

$$\sum_{j=1}^{r} q_{ij} = 1 \qquad \text{pre každé } i = 1, 2, \ldots, r.$$

Hovoríme, že množina indexov $\mathcal{I} \subseteq \{1, 2, \ldots, r\}$ je **uzavretá**, ak

$$\sum_{j \in I} q_{ij} = 1 \qquad \text{pre každé } i \in \mathcal{I}.$$

Nech $\mathbf{\Pi} = (q_{ij})$ je stochastická matica typu $r \times r$. Hovoríme, že matica $\mathbf{\Pi}$ je **rozložiteľná**, ak existuje uzavretá množina indexov $\mathcal{I} \subset \{1, 2, \ldots, r\}$ taká, že $\mathcal{I} \neq \{1, 2, \ldots, r\}$.
Hovoríme, že matica $\mathbf{\Pi}$ je **nerozložiteľná**, ak jedinou uzavretou množinou indexov je množina $\{1, 2, \ldots, r\}$ všetkých indexov.

Príklad 3.3. Nech $\mathbf{\Pi} = (q_{ij})$ je stochastická matica typu $r \times r$, ktorej riadky a stĺpce zodpovedajú prvkom abecedy Z. Prvok $q_{ij} = P(X_2 = a_j | X_1 = a_i)$ je podmienená pravdepodobnosť javu, že zdroj v čase 2 vyšle znak a_j za predpokladu, že v čase 1 vyslal znak a_i.

Nech $\boldsymbol{p} = (p_1, p_2, \ldots, p_r)$ je taký riadkový vektor, pre ktorý je $\boldsymbol{p}.\boldsymbol{\Pi} = \boldsymbol{p}$. Na maticu $\boldsymbol{\Pi}$ nekladieme žiadne ďalšie predpoklady. Definujme

$$P(x_1, x_2, \ldots, x_k) = p_{x_1}.q_{x_1x_2}.q_{x_2x_3}.\ldots.q_{x_{k-1}x_k} \tag{3.32}$$

Ľahko overíme, že funkcie $P_k()$ definované v (3.32) spĺňajú (3.28), (3.29), (3.30) a pretože $\boldsymbol{p}.\boldsymbol{\Pi} = \boldsymbol{p}$, platí aj (3.31). Existuje teda miera P na $\mathcal{F}$, pre ktorú platí (3.27). Zobrazenie T na priestore $\mathcal{Z} = (\Omega, \mathcal{F}, P, T)$ nazveme Markovským posunom, zdroj $\mathcal{Z}$ nazveme Markovským zdrojom.

Označme $q_{ij}^{(k)}$ pravdepodobnosť, že zdroj po vyslaní znaku a_i po k krokoch vyšle znak a_j, t. j. $P\{X_1 = a_i, X_{k+1} = a_j\} = p_i.q_{ij}^{(k)}$, $q_{ij}^{(0)} = 1$, ak $i = j$, inak $q_{ij}^{(0)} = 0$. Ďalej označme

$$s_{ij} = \lim_{n\to\infty} \frac{1}{n}.\sum_{k=0}^{n-1} q_{ij}^{(k)}$$

Pre takto definovaný zdroj platí nasledujúca veta (pozri [3]), ktorú uvedieme bez dôkazu.

Veta 3.12. *Nasledujúce štyri tvrdenia sú ekvivalentné*

a) Zobrazenie T je ergodické.

b) Veličiny s_{ij} nezávisia od i.

c) Matica $\boldsymbol{\Pi}$ je nerozložiteľná.

d) Pre ľubovoľné $i,\ j$ je $s_{ij} > 0$.

Ergodicita zdroja je veľmi silná vlastnosť. Pre ergodické zdroje vždy existuje entropia. Pre ergodické zdroje platí Shannonova – Mac Millanova veta (ktorú sme doteraz formulovali len pre stacionárny nezávislý zdroj).

Ako sme už ukázali, písaný jazyk (napr. slovenčina) je síce s istým priblížením stacionárnym zdrojom, ale ani zďaleka nie je nezávislým zdrojom. Ak javy $A,\ B$ sú dve slová (t. j. elementárne cylindre), potom $T^{-n}B$ s veľkým n predstavuje jav, že slovo B bude vyslané v ďalekej budúcnosti. Dá sa predpokladať, že čím väčšie bude n, a teda čím väčší bude časový interval medzi vyslaním slova A a slova $T^{-n}B$, tým menej bude jav $T^{-n}B$ závisieť od javu A. Môžeme teda predpokladať, že $\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A).P(B)$, a teda že zobrazenie T je premiešavajúce, z čoho vyplýva ergodicita zobrazenia T. Písaný jazyk môžeme teda s dobrým priblížením považovať za ergodický zdroj informácie, a teda za

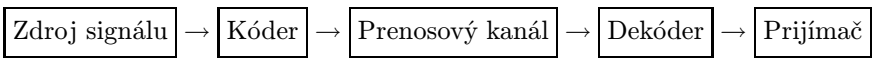
zdroj pre ktorý platí Shannonova–Mac Millanova veta a mnohé ďalšie tvrdenia, z ktorých z najdôležitejšie možno považovať priamu a obrátenú Shannonovu vetu (vety 5.1 a 5.2 na str. 146).

Kapitola 4

Kódovanie

4.1 Prenosový reťazec

Všeobecná schéma prenosového reťazca je nasledujúca:



Zdroj signálu, prenosový kanál a prijímacie zariadenie môžu pracovať v úplne iných abecedách. Vysielacie štúdio má hudobnú skladbu uloženú na CD nosiči – teda v binárnom kóde. Aby ju mohlo vysielať FM prenosom, musí z tohoto binárneho kódu vytvoriť vysokofrekvenčný signál (okolo 100 MHz, čo je signál, ktorým pracuje prenosový kanál), čo nie je nič iného, ako zakódovanie. Rádiový prijímač prijme FM signál a spracuje ho do konečnej podoby zvukových vĺn - dekóduje signál kanálu.

Ak chceme preniesť správu pomocou baterky schopnej vydávať len signály bodka, čiarka a medzipísmenová medzera, musíme prenášaný text upraviť - zakódovať pomocou napr. Morseovej abecedy do týchto znakov.

Hlavným dôvodom kódovania správ je prispôsobenie abecedy zdroja abecede prenosového kanála. Pritom však môžeme mať i ďalšie ciele – napríklad chceme, aby zakódovaná správa bola čo najkratšia (kompresia správ). Na druhej strane môžeme žiadať, aby bolo možné poznať, či v zakódovanej správe nedošlo behom prenosu k jednej alebo viacerým chybám, resp. aby zakódovaná správa bola odolná voči jednej či viacerým chybám pri prenose. Navyše chceme, aby kompresia, zisťovanie chyby, či odstraňovanie chyby boli výpočtovo čo najjednoduchšie. Jednotlivé požiadavky sú protichodné a nie je vždy jednoduché zaistiť

ich samotné, nie to ešte v kombináciách. Kódovanie nemá za cieľ utajenie správy, preto je nutné rozlišovať ho od šifrovania – kryptografie, ktorých cieľom ochrana a utajenie dát.

Problémami kódovania, kompresie dát, kódov odhaľujúcich chyby a samoopravných kódov sa zaoberá teória kódovania, ktorej základy uvedieme v tejto kapitole.

4.2 Abeceda, kód a kódovanie

Nech $A = \{a_1, a_2, \ldots, a_r\}$ je konečná r-prvková množina. Prvky množiny A nazveme **znakmi**, množinu A **abecedou**. Množinu

$$A^* = \bigcup_{i=1}^{\infty} A^i$$

nazveme **množinou všetkých slov abecedy** A. **Dĺžka slova** $\mathbf{a} \in A^*$ je počet znakov slova **a**.
Na množine slov abecedy A zavádzame binárnu **operáciu zreťazenia slov**: Ak $\mathbf{b} = b_1 b_2 \ldots b_p$, $\mathbf{c} = c_1 c_2 \ldots c_q$ sú dve slová z A^*, potom definujeme

$$\mathbf{b}|\mathbf{c} = b_1 b_2 \ldots b_p c_1 c_2 \ldots c_q.$$

Zreťazenia slov píšeme bez medzery, či iného oddeľovacieho znaku. Každé slovo môžeme považovať za zreťazenie jeho častí ľubovoľným spôsobom, ako sa nám hodí. Tak napríklad $01010001 = 0101|0001 = 010|100|01 = 0|1|0|1|0|0|0|1$.

Nech $A = \{a_1, a_2, \ldots, a_r\}$, $B = \{b_1, b_2, \ldots, b_s\}$ sú dve abecedy. **Kódovanie** je zobrazenie

$$K : A \to B^*,$$

t. j. predpis, ktorý každému prvku abecedy A priradí slovo abecedy B. Abeceda A je **zdrojová abeceda**, jej znaky sú **zdrojové znaky**, abeceda B je **kódová abeceda** a jej znaky sú **kódové znaky**. Množinu všetkých slov kódovej abecedy typu

$$\mathcal{K} = \{\mathbf{b} \mid \mathbf{b} = K(a),\ a \in A\} = \{K(a_1), K(a_2), \ldots, K(a_r)\}$$

nazveme **kódom**, každé slovo z množiny $\mathcal{K}$ je **kódové slovo** ostatné slová z abecedy B sú **nekódové slová**. Význam majú iba prosté kódovania, t. j. také, kde rôznym zdrojovým znakom $a_i,\ a_j \in A$ zodpovedajú rôzne kódové slová

$K(a_i)$, $K(a_j)$, preto budeme vždy predpokladať, že zobrazenie K je prosté.
Každé kódovanie K môžeme rozšíriť na kódovanie K^* zdrojových slov predpisom

$$K^*(a_{i_1}a_{i_2}\ldots_{i_n}) = K(a_{i_1})|K(a_{i_2})|\ldots|K(a_{i_n})$$

Kódovanie K^* je vlastne kódovaním znak po znaku.
Kódovanie môže rôznym znakom priradiť kódové slová rôznej dĺžky. Často sa však stretávame s kódovaniami, u ktorých všetky kódové slová majú rovnakú dĺžku. **Blokové kódovanie** (dĺžky n) je také kódovanie, ktoré všetkým zdrojovým znakom priradí kódové slová rovnakej dĺžky n.

Príklad 4.1. Nech $A = \{a, b, c, d\}$, $B = \{0, 1\}$, nech $K(a) = 00$, $K(b) = 01$, $K(c) = 10$, $K(d) = 11$. Potom správu $aabd$ (t. j. slovo v abecede A) zakódujeme ako $K^*(aabd) = 00000111$. Ak na strane prijímača dostaneme slovo 00000111 a poznáme zobrazenie K, vieme, že každý znak zdrojovej abecedy bol zakódovaný do dvoch znakov kódovej abecedy, a teda jediné možné rozdelenie prijatej správy na kódové slová je 00|00|01|11, čo vedie k jednoznačnému dekódovaniu správy. Kódovanie K je blokovým kódovaním dĺžky 2.

Príklad 4.2. Študenti sú hodnotení známkami 1, 2, 3, 4. Vieme, že najčastejšia známka je 2 a potom 1. Na zakódovanie štyroch znakov zdrojovej abecedy $A = \{1, 2, 3, 4\}$ by stačili dva znaky binárnej kódovej abecedy $B = \{0, 1\}$. Pretože však trojky a štvorky sa vyskytujú zriedkavo, a dvojky zas veľmi často, chceme dvojkám dať čo najkratšie kódové slovo. Navrhneme preto toto kódovanie: $K(1) = 01$, $K(2) = 0$, $K(3) = 011$, $K(4) = 111$. Správa 1234 bude zakódovaná ako 01|0|011|111. Ak budeme postavení pred úlohu dekódovať správu 010011111, budeme musieť postupovať od zadu. Ak napríklad dostaneme čiastočnú správu $01111\ldots$, nevieme, či bola vyslaná ako $0|111|1\ldots$, alebo $01|111\ldots$, alebo $011|11\ldots$, nemôžeme ho preto dekódovať znak po znaku.

Definícia 4.1. Hovoríme, že kódovanie $K : A \to B^*$ je **jednoznačne dekódovateľné**, ak zo znalosti zakódovanej správy $K^*(a_1a_1\ldots a_n)$ môžeme vždy určiť zdrojovú správu $a_1a_1\ldots a_n$, t. j. ak je zobrazenie $K^* : A^* \to B^*$ prostým zobrazením.

Príklad 4.3. Rozšírme zdrojovú abecedu z príkladu 4.2 na $A = \{1, 2, 3, 4, 5\}$ a definujme kódovanie $K(1) = 01$, $K(2) = 0$, $K(3) = 011$, $K(4) = 111$, $K(5) = 101$. Majme správu 0101101. Pre dekódovanie by sme ju mohli rozdeliť nasledovne: 0|101|101, 01|01|101, 01|011|01, pričom tieto delenia zodpovedajú zdrojovým slovám porade 255, 115, 131. Vidíme, že napriek tomu, že kódové zobrazenie $K : A \to B^*$ je prosté, príslušné zobrazenie $K^* : A^* \to B^*$ prosté nie je. K nie je jednoznačne dekódovateľné kódovanie.

4.3 Prefixové kódovanie a Kraftova nerovnosť

Definícia 4.2. Prefixom slova b $= b_1 b_2 \dots b_k$ nazveme každé zo slov b_1, $b_1 b_2, \dots$, $b_1 b_2 \dots b_{k-1}$, $b_1 b_2 \dots b_k$. Kódovanie resp. kód sa nazýva **prefixové**, ak žiadne kódové slovo nie je prefixom iného kódového slova.

Všimnime si, že každé blokové kódovanie je prefixovým kódovaním. Rozšíreným prefixovým kódovaním, s ktorým sa všetci dennodenne stretávame, je priradenie telefónnych čísel staniciam v sieti Slovenských telekomunikácií. Telefónne čísla nie sú blokovým kódom, pretože čísla staníc majú rôznu dĺžku – napr, 120 – informácie resp. 155 – záchranná služba sú len trojmiestne, ale pre všetky stanice s troma dekadickými číslicami nevystačíme. Číslo žiadnej telefónnej stanice nemôže začínať číslom inej stanice, pretože by sa v priebehu vytáčania dlhšieho čísla vždy ohlásila stanica s prefixom. Tak napríklad číslo 120 je vyhradené pre informácie, a žiadne iné telefónne číslo v sieti Slovenských telekomunikácií nemá číslo typu 120XXX.

Prefixové kódovanie je jediné kódovanie, ktoré môžeme dekódovať znak po znaku – t. j. v priebehu prijímania správy (a nemusíme čakať na prijatie celej správy). Dekódovanie prijatej správy robíme tak, že v nej nájdeme najmenší počet znakov zľava, ktoré tvoria kódové slovo $K(a)$ niektorého zdrojového znaku a, tieto znaky dekódujeme, zrušíme dekódované znaky z kódovanej správy a pokračujeme ďalej rovnakým spôsobom.

Veta 4.1. Kraftova nerovnosť. *Majme zdrojovú abecedu* $A = \{a_1, a_2, \dots, a_r\}$ *s* r *znakmi, kódovú abecedu* $B = \{b_1, b_2, \dots, b_n\}$ *s* n *znakmi. Prefixový kód s dĺžkami kódových slov* $d_1, d_2, \dots, d_r$ *existuje práve vtedy, keď*

$$n^{-d_1} + n^{-d_2} + \cdots + n^{-d_r} \leq 1. \tag{4.1}$$

Dôkaz. Nech platí Kraftova nerovnosť (4.1). Usporiadajme znaky zdrojovej abecedy tak, aby platilo $d_1 \leq d_2 \leq \cdots \leq d_r$. Za $K(a_1)$ zvoľme ľubovoľné slovo abecedy B dĺžky d_1. Predpokladajme, že už máme priradené kódové slová požadovanej dĺžky $K(a_1)$, $K(a_2)$, $\dots$, $K(a_i)$. Pri voľbe kódového slova $K(a_{i+1})$ dĺžky d_{i+1} sa musíme vyhnúť $n^{(d_{i+1}-d_1)}$ slovám dĺžky d_{i+1}, ktoré majú prefix $K(a_i)$, $n^{(d_{i+1}-d_2)}$ slovám dĺžky d_{i+1}, ktoré majú prefix $K(a_{i+1})$ atď. až $n^{(d_{i+1}-d_i)}$ slovám dĺžky d_{i+1}, ktoré majú prefix $K(a_i)$, pričom všetkých slov dĺžky d_{i+1} je $n^{d_{i+1}}$. Počet zakázaných slov je teda

$$n^{(d_{i+1}-d_1)} + n^{(d_{i+1}-d_2)} + \cdots + n^{(d_{i+1}-d_i)}. \tag{4.2}$$

Keďže platí (4.1), tým skôr platí pre prvých $i+1$ členov ľavej strany (4.1):

$$n^{-d_1} + n^{-d_2} + \cdots + n^{-d_i} + n^{-d_{i+1}} \leq 1. \tag{4.3}$$

Po vynásobení nerovnosti (4.3) číslom $n^{d_{i+1}}$ dostávame

$$n^{(d_{i+1}-d_1)} + n^{(d_{i+1}-d_2)} + \cdots + n^{(d_{i+1}-d_i)} + 1 \leq n^{d_{i+1}}. \tag{4.4}$$

Podľa (4.4) je počet zakázaných slov aspoň o 1 slovo menší, než počet všetkých slov dĺžky d_{i+1} a preto môžeme toto slovo definovať ako kódové slovo $K(a_{i+1})$.

Majme prefixový kód s dĺžkami $d_1, d_2, \ldots, d_r$. Predpokladajme $d_1 \leq d_2 \leq \cdots \leq d_r$. Existuje n^{d_r} slov dĺžky d_r, ktorými možno zakódovať písmeno a_r. Pre každé $i = 1, 2, \ldots, r-1$ je slovo $K(a_i)$ prefixom $n^{(d_r-d_i)}$ slov dĺžky d_r – tieto slová sú pre výber slova $K(a_r)$ zakázané (inak by totiž kód nebol prefixový). Pretože aj pre slovo $K(a_r)$ sa ušlo jedno kódové slovo dĺžky d_r, musí platiť:

$$n^{(d_r-d_1)} + n^{(d_r-d_2)} + \cdots + n^{(d_r-d_{r-1})} + 1 \leq n^{d_r} \ . \tag{4.5}$$

Vydelením nerovnosti (4.5) číslom n^{d_r} dostávame požadovanú Kraftovu nerovnosť (4.1). ■

Poznámka. **Algoritmus na zostrojenie prefixového kódu s dĺžkami slov $d_1, d_2, \ldots, d_r$.** Prvá časť dôkazu vety 4.1 je konštruktívna, dáva návod na zostrojenie prefixového kódovania, ak sú dané požadované dĺžky $d_1 \leq d_2 \leq \cdots \leq d_r$ kódových slov spĺňajúce Kraftovu nerovnosť. Za $K(a_1)$ zvolíme ľubovoľné slovo dĺžky d_1. Keď už máme určené $K(a_1)$, $K(a_2)$, $\ldots$, $K(a_i)$, za $K(a_{i+1})$ zvolíme ľubovoľné slovo dĺžky d_{i+1}, ktoré nemá ako prefix žiadne zo slov $K(a_1), K(a_2), \ldots, K(a_i)$. Existenciu aspoň jedného takéhoto slova zaručuje Kraftova nerovnosť.

Veta 4.2. Mac Millan.
Pre každé jednoznačne dekódovateľné kódovanie so zdrojovou abecedou $A = \{a_1, a_2, \ldots, a_r\}$ a kódovou abecedou $B = \{b_1, b_2, \ldots, b_n\}$ s dĺžkami kódových slov $d_1, d_2, \ldots, d_r$ platí Kraftova nerovnosť (4.1).

Dôkaz. Majme jednoznačne dekódovateľné kódovanie K s dĺžkami kódových slov $d_1 \leq d_2 \leq \cdots \leq d_r$. Označme

$$c = n^{-d_1} + n^{-d_2} + \cdots + n^{-d_r} \ . \tag{4.6}$$

V ďalšom postupe sa budeme snažiť ukázať, že $c \leq 1$.
Nech k je ľubovoľné prirodzené číslo. Majme množinu $\mathcal{M}_k$ všetkých slov kódovej abecedy typu $\mathbf{b} = K(a_{i_1})|K(a_{i_2})|\ldots|K(a_{i_k})$. Dĺžka každého takéhoto slova $\mathbf{b}$ je $d_{i_1} + d_{i_2} + \cdots + d_{i_k}$ a je menšia alebo rovná $k.d_r$, pretože maximálna dĺžka

kódového slova je d_r.
Skúmajme výraz

$$c^k = \left[n^{-d_1} + n^{-d_2} + \cdots + n^{-d_r}\right]^k = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} n^{-(d_{i_1}+d_{i_2}+\cdots+d_{i_k})} \ . \quad (4.7)$$

Pretože K je jednoznačne dekódovateľné, platí pre dve rôzne slová zdrojovej abecedy $a_1 a_2 \ldots a_{i_k}$, $a'_1 a'_2 \ldots a'_{i_k}$

$$K(a_{i_1})|K(a_{i_2})|\ldots|K(a_{i_k}) \neq K(a'_{i_1})|K(a'_{i_2})|\ldots|K(a'_{i_k}) \ .$$

Preto ku každému slovu $\mathbf{b} = K(a_{i_1})|K(a_{i_2})|\ldots|K(a_{i_k})$ z množiny $\mathcal{M}_k$ možno priradiť práve jeden sčítanec $n^{-(d_{i_1}+d_{i_2}+\cdots+d_{i_k})}$ na pravej strane (4.7) taký, že jeho záporne vzatý exponent $(d_{i_1}+d_{i_2}+\cdots+d_{i_k})$ sa rovná dĺžke slova $\mathbf{b}$. Ako sme už ukázali, maximálna dĺžka slova z množiny $\mathcal{M}_k$ je kd_r. Označme $M = kd_r$. Výraz na pravej strane vzťahu (4.7) je polynómom stupňa M premennej $\frac{1}{n}$, a preto ho môžeme zapísať v tvare

$$c^k = s_1.n^{-1} + s_2.n^{-2} + \cdots + s_M.n^{-M} = \sum_{i=1}^{M} s_i.n^{-i}.$$

V súčte na pravej strane (4.7) sa vyskytuje člen n^{-i} práve toľkokrát, koľko slov z množiny $\mathcal{M}_k$ má dĺžku i. Pretože kódová abeceda má n znakov, najviac n^i slov z množiny $\mathcal{M}_k$ môže mať dĺžku i, čo znamená, že $s_i \leq n^i$. Môžeme teda písať:

$$\begin{aligned} c^k &= s_1.n^{-1} + s_2.n^{-2} + \cdots + s_M.n^{-M} \leq \\ &\leq n^1.n^{-1} + n^2.n^{-2} + \cdots + n^M.n^{-M} \leq 1 + 1 + \cdots + 1 = M = k.d_r \end{aligned} \quad (4.8)$$

a teda

$$\frac{c^k}{k} \leq d_r \ . \quad (4.9)$$

Pretože nerovnosť (4.9) musí platiť pre ľubovoľné k, musí byť $c \leq 1$. ■

Dôsledkom Mac Millanovej vety je, že žiadnym jednoznačne dekódovateľným kódovaním nedosiahneme kratšie dĺžky kódových slov ako prefixovým kódovaním. Keďže prefixové kódovanie má veľa výhod – jednoduché dekódovanie znak po znaku, netreba čakať na príjem celej správy – stačí sa obmedziť na prefixové kódovanie.

4.4 Najkratší kód - Huffmanova konštrukcia

Nech je daný stacionárny zdroj $\mathcal{Z} = (\Omega, \mathcal{A}, P)$, ktorý produkuje jednotlivé znaky zdrojovej abecedy $A = \{a_1, a_2, \ldots, a_r\}$ s pravdepodobnosťami $p_1, p_2, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$. Majme prefixové kódovanie K také, že dĺžky kódových slov $K(a_1)$, $K(a_2)$, $\ldots$, $K(a_r)$ sú $d_1, d_2, \ldots, d_r$. Potom **stredná dĺžka kódového slova** kódovania K je

$$d(K) = p_1.d_1 + p_2.d_2 + \cdots + p_r.d_r = \sum_{i=1}^{r} p_i.d_i \ . \tag{4.10}$$

Ak kódovaním K kódujeme správu s veľkým počtom N znakov, môžeme očakávať, že dĺžka (počet znakov) zakódovanej správy v abecede B bude približne $N.d(K)$. Keďže veľmi často (z hľadiska prenosu alebo uloženia správy) chceme, aby zakódovaná správa bola čo najkratšia, hľadáme kódovanie K s minimálnou strednou dĺžkou kódového slova $d(K)$.

Definícia 4.3. Majme danú zdrojovú abecedu $A = \{a_1, a_2, \ldots, a_r\}$ s pravdepodobnosťami výskytu $p_1, p_2, \ldots, p_r$ a kódovú abecedu $B = \{b_1, b_2, \ldots, b_n\}$. **Najkratšie n-znakové kódovanie** abecedy A je také kódovanie $K : A \to B^*$, ktoré má najmenšiu strednú dĺžku kódového slova $d(K)$.

Najkratší prefixový kód skonštruoval O. Huffman (čítaj hafmen) v roku 1952. Budeme sa zaoberať hlavne binárnym kódovaním, pretože je z hľadiska aplikácií najdôležitejšie. Predpokladáme, že zdrojové znaky $a_1, a_2, \ldots, a_r$ sú usporiadané zostupne podľa pravdepodobnosti, t. j. $p_1 \geq p_2 \geq \cdots \geq p_r$.

Ak máme len dva znaky a_1, a_2 s ľubovoľnými pravdepodobnosťami, je situácia jednoduchá – najkratšie kódovanie je $K(a_1) = 0$, $K(a_2) = 1$ (alebo $K(a_1) = 1$, $K(a_2) = 0$).

Definícia 4.4. Majme abecedu s r znakmi $A = \{a_1, a_2, \ldots, a_r\}$ s pravdepodobnosťami znakov $p_1 \geq p_2 \geq \cdots \geq p_r$. **Redukovanou abecedou** abecedy A nazveme abecedu $\tilde{A} = \{\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{r-1}\}$, kde $\tilde{a}_i = a_i$ pre $i = 1, 2 \ldots, r-2$, $\tilde{a}_{r-1} = a^*$, kde $a^* \neq A$ a $p(\tilde{a}_i) = p(a_i)$ pre $i = 1, 2 \ldots, r-2$, $p(\tilde{a}_{r-1}) = p_{r-1} + p_{r-2}$.

Veta 4.3. *Nech $\tilde{A}$ je redukovaná abeceda abecedy A, nech $\tilde{K}$ je najkratšie prefixové binárne kódovanie redukovanej abecedy $\tilde{A}$. Potom kódovanie K definované*

$$K(a_1) = \tilde{K}(\tilde{a}_1), \quad K(a_2) = \tilde{K}(\tilde{a}_2), \quad \ldots, \quad K(a_{r-2}) = \tilde{K}(\tilde{a}_{r-2}), \tag{4.11}$$

$$K(a_{r-1}) = \tilde{K}(\tilde{a}_{r-1})|0, \quad K(a_r) = \tilde{K}(\tilde{a}_r)|1, \tag{4.12}$$

je najkratším prefixovým binárnym kódovaním abecedy A.

Dôkaz. Dokážeme najprv dve pomocné tvrdenia.

Lema 4.1. *Nech $A = \{a_1, a_2, \ldots, a_r\}$ je zdrojová abeceda s pravdepodobnosťami výskytu znakov $p_1 \geq p_2 \geq \cdots \geq p_r$. Potom možno zostrojiť najkratší prefixový binárny kód K'' taký, že kódové slová $K''(a_{r-1})$, $K''(a_r)$ sa líšia len v poslednom znaku.*

Dôkaz. Nech K' je najkratšie prefixové binárne kódovanie abecedy A. Ak označíme d'_i dĺžku slova $K'(a_i)$, musí platiť

$$d'_1 \leq d'_2 \leq \cdots \leq d'_r \ . \tag{4.13}$$

Ak by totiž $d'_i > d'_{i+1}$, vzájomnou zámenou $K'(a_i)$ a $K'(a_{i+1})$ by sme dostali prefixové kódovanie s menšou strednou dĺžkou kódového slova.

Vytvorme nové kódovanie K^* tak, že $K^*(a_i) = K'(a_i)$ pre všetky $i = 1, 2, \ldots, r-1$ a $K^*(a_r)$ bude to slovo, ktoré vznikne zo slova $K'(a_r)$ vynechaním posledného znaku. Kódovanie K^* vzniklo z najkratšieho prefixového kódovania K', samo však už nemôže byť prefixovým kódovaním, lebo by malo menšiu strednú dĺžku kódového slova než kódovanie K'. Slovo $K^*(a_r)$ musí byť preto prefixom nejakého iného slova $K'(a_j)$. (Je to totiž jediné skrátené slovo prefixového kódovania K'). Preto $d'_r - 1 < d'_j$ resp. $d'_r \leq d'_j$, avšak vzhľadom na (4.13) $d'_j \leq d'_r$, z čoho máme $d'_j = d'_r$. Preto

$$d'_1 \leq d'_2 \leq \ldots d'_j = d'_{j+1} = \cdots = d', \tag{4.14}$$

t. j. všetky kódové slová začínajúc slovom $K'(a_j)$ až po posledné slovo $K'(a_r)$ majú rovnakú dĺžku, pričom slová $K'(a_j)$, $K'(a_r)$ sa líšia len vo svojom poslednom znaku. Ak sú tieto dve slová posledné, t. j. ak $j = r-1$, K' je hľadané kódovanie K''. Inak z kódovania K' zostrojíme kódovanie K'' tak, že vzájomne zameníme kódy znakov a_j, a_{r-1}. Exaktne zapísané $K''(a_i) = K'(a_i)$ pre všetky $i = 1, 2, \ldots, r$, $i \neq j$, $i \neq r-1$, $K''(a_j) = K'(a_{r-1})$, $K''(a_{r-1}) = K'(a_j)$. ■

Lema 4.2. *Nech $\tilde{K}$ je prefixový kód redukovanej abecedy, nech K je kód pôvodnej abecedy $A = \{a_1, a_2, \ldots, a_r\}$ s pravdepodobnosťami $p_1, p_2, \ldots, p_r$, pre ktorý platí* (4.11), (4.12). *Označme d_i resp. $\tilde{d}_i$ dĺžku slova $K(a_i)$ resp. $\tilde{K}(a_i)$ a $d(K)$ resp. $d(\tilde{K})$ stredné dĺžky kódových slov kódovaní K, $\tilde{K}$. Potom platí:*

$$d(K) - d(\tilde{K}) = p_{r-1} + p_r \ .$$

Dôkaz. Pretože platí (4.11) je $d_i = \tilde{d}_i$ pre $i = 1, 2, \ldots r-2$. Pretože (4.12), $d_{r-1} = d_r = \tilde{d}_{r-1} + 1$. Pre pravdepodobnosti znakov redukovanej abecedy platí $\tilde{p}_i = p_i$ pre $i = 1, 2, \ldots r-2$, $\tilde{p}_{r-1} = p_{r-1} + p_r$. S využitím týchto vzťahov môžeme písať

$$\begin{aligned} d(\tilde{K}) &= \sum_{i=1}^{r-1} \tilde{p}_i.\tilde{d}_i = p_1.d_1 + p_2.d_2 + \cdots + p_{r-2}.d_{r-2} + (p_{r-1} + p_r).\tilde{d}_{r-1} \ , \\ d(K) &= \sum_{i=1}^{r} p_i.d_i = p_1.d_1 + p_2.d_2 + \cdots + p_{r-2}.d_{r-2} + \\ &\qquad + p_{r-1}.(\tilde{d}_{r-1} + 1) + p_r.(\tilde{d}_{r-1} + 1) \ , \end{aligned}$$

odkiaľ máme

$$d(K) - d(\tilde{K}) = (p_{r-1} + p_r).(\tilde{d}_{r-1} + 1) - (p_{r-1} + p_r).\tilde{d}_{r-1} = p_{r-1} + p_r \ .$$

■

Konečne dokážeme tvrdenie vety. Nech $\tilde{K}$ je najkratší prefixový kód redukovanej abecedy. Nech kód K je definovaný pomocou vzťahov (4.11), (4.12). Aj kód K je prefixový (lebo $\tilde{K}$ bol prefixový). Podľa lemy 4.1 možno zostrojiť najkratší prefixový kód K'' taký, že kódové slová $K''(a_{r-1})$, $K''(a_r)$ sa líšia len v poslednom znaku. Ku kódu K'' možno jednoznačne zostrojiť kód $\tilde{K}''$ kód redukovanej abecedy, takže pre oba kódy platia vzťahy (4.11), (4.12), kde namiesto K píšeme K''. Preto podľa lemy 4.2 platí

$$d(K'') - d(\tilde{K}'') = p_{r-1} + p_r \ , \qquad \text{čiže} \quad d(K'') = d(\tilde{K}'') + p_{r-1} + p_r \ .$$

Z tých istých dôvodov môžeme písať

$$d(K) - d(\tilde{K}) = p_{r-1} + p_r \ , \qquad \text{čiže} \quad d(K) = d(\tilde{K}) + p_{r-1} + p_r \ .$$

Pretože však $\tilde{K}$ bol najkratší kód redukovanej abecedy, musí byť $d(\tilde{K}) \leq d(\tilde{K}'')$ a preto

$$d(K) = d(\tilde{K}) + p_{r-1} + p_r \leq d(\tilde{K}'') + p_{r-1} + p_r = d(K'') \ .$$

Pretože K'' bol najkratší prefixový kód, je

$$d(K'') \leq d(K).$$

Z posledných dvoch nerovností máme $d(K) = d(K'')$ – aj kódovanie K je najkratším binárnym kódovaním. ■

4.5 Algoritmus na zostrojenie Huffmanovho kódu

Budeme postupne budovať binárny koreňový strom, ktorého listy budú znaky zdrojovej abecedy A. Každý vrchol stromu bude mať priradenú pravdepodobnosť a binárny znak 0 alebo 1

Krok 1: Zostroj hranovo ohodnotený graf $G = (V, H, p)$, kde $V = A$ a kde $p(v)$ je pravdepodobnosť znaku v. Inicializačne polož $H := \emptyset$. Všetky vrcholy z V inicializačne prehlás za neoznačené.

Krok 2: Nájdi dva neoznačené vrcholy u, w z množiny V s najmenšími pravdepodobnosťami $p(u)$, $p(w)$. Označkuj vrchol u značkou 0, vrchol w značkou 1. Množinu vrcholov V rozšír o vrchol x, t. j. polož $V := V \cup \{x\}$ pre nejaké $x \notin V$, polož $p(x) := p(u) + p(w)$, $H := H \cup \{(x, u), (x, w)\}$ a nový vrchol x prehlás za neoznačený.

Krok 3: Ak je graf G súvislý, choď na Krok 4, inak pokračuj Krokom 2.

Krok 4: Teraz je graf G koreňovým stromom s listami (t. j. vrcholmi stupňa 1) zodpovedajúcimi znakom zdrojovej abecedy A. Všetky vrcholy stromu G okrem koreňa sú označené binárnymi značkami 0 alebo 1. Z koreňa stromu do každého listu vedie jediná cesta, postupnosť binárnych značiek vrcholov na tejto ceste určuje prefixový kód príslušného znaku.

Analogicky sa skonštruuje i n-árny Huffmanov kód. Predpokladajme, že abeceda A má $r = n + k.(n - 1)$ znakov. Keby nie doplníme abecedu A o fiktívne znaky s nulovou pravdepodobnosťou – kódové slová priradené týmto fiktívnym znakom zostanú nevyužité. Nájdeme n znakov zdrojovej abecedy s najmenšími pravdepodobnosťami a týmto znakom priradíme znaky kódovej abecedy v ľubovoľnom poradí – budú to posledné znaky ich kódových slov. Abecedu A zredukujeme tak, že namiesto n znakov s najmenšími pravdepodobnosťami

dodáme jeden znak so súčtom pravdepodobností nahradených znakov. Zredukovaná abeceda má $n + (k-1).(n-1)$ znakov. Ak $k - 1 > 0$ urobíme znovu redukciu atď.

4.6 Vzťah entropie zdroja a dĺžky najkratšieho kódovania

V definícii 3.5 (str. 54) bola definovaná entropia zdroja informácie ako

$$H(\mathcal{Z}) = -\lim_{n\to\infty} \frac{1}{n} . \sum_{(x_1,\ldots,x_n)\in Z} P(x_1, x_2, \ldots, x_n).\log_2 P(x_1, x_2, \ldots, x_n) \ . \tag{4.15}$$

Pre stacionárny nezávislý zdroj $\mathcal{Z}$ s r-prvkovou abecedou $A = \{a_1, a_2, \ldots, a_r\}$ s pravdepodobnosťami znakov $p_1, p_2, \ldots, p_r$ sme ukázali, (veta 3.1 na str. 55) že

$$H(\mathcal{Z}) = -\sum_{i=1}^{r} p_1.\log_2(p_i) \ .$$

Majme ľubovoľné binárne prefixové kódovanie K abecedy A s dĺžkami kódových slov $d_1, d_2, \ldots, d_r$ a so strednou dĺžkou slova $d = d(K)$. Chceme zistiť vzťah medzi veličinami $H(\mathcal{Z})$ a d pre prípad stacionárneho nezávislého zdroja. Môžeme postupne písať:

$$\begin{aligned}
H(\mathcal{Z}) - d &= \sum_{i=1}^{r} p_i.\log_2\left(\frac{1}{p_i}\right) - \sum_{i=1}^{r} p_i.d_i = \sum_{i=1}^{r} p_i.\left[\log_2\left(\frac{1}{p_i}\right) - d_i\right] = \\
&= \sum_{i=1}^{r} p_i.\left[\log_2\left(\frac{1}{p_i}\right) + \log_2\left(2^{-d_i}\right)\right] = \sum_{i=1}^{r} p_i.\left[\log_2\left(\frac{2^{-d_i}}{p_i}\right)\right] = \\
&= \frac{1}{\ln 2}.\sum_{i=1}^{r} p_i.\left[\ln\left(\frac{2^{-d_i}}{p_i}\right)\right] \leq \\
&\leq \frac{1}{\ln 2}.\sum_{i=1}^{r} p_i.\left(\frac{2^{-d_i}}{p_i} - 1\right) = \frac{1}{\ln 2}.\left[\sum_{i=1}^{r} 2^{-d_i} - \sum_{i=1}^{r} p_i\right] = \\
&= \frac{1}{\ln 2}.\left[\sum_{i=1}^{r} 2^{-d_i} - 1\right] \leq 0 \ .
\end{aligned}$$

Prvá nerovnosť v predchádzajúcom postupe vyplýva z nerovnosti $\ln(x) \leq x - 1$ aplikovanej na $x = 2^{-d_i}/p_i$, druhá nerovnosť platí preto, lebo prirodzené čísla d_i sú dĺžkami kódových slov prefixového kódovania a platí pre ne Kraftova nerovnosť $\sum_{i=1}^{r} 2^{-d_i} \leq 1$. Platí teda $H(\mathcal{Z}) \leq d(K)$ pre každé prefixové (a teda aj jednoznačne dekódovateľné) kódovanie.
Zvoľme teraz prirodzené čísla d_i pre $i = 1, 2, \ldots, r$ tak, aby platilo

$$-\log_2(p_i) \leq d_i < -\log_2(p_i) + 1$$

pre každé i.
Potom prvú nerovnosť môžeme postupne prepísať

$$\log_2\left(\frac{1}{p_i}\right) \leq d_i \quad \Rightarrow \quad \frac{1}{p_i} \leq 2^{d_i} \quad \Rightarrow \quad 2^{-d_i} \leq p_i \ .$$

Keďže posledná nerovnosť v predchádzajúcom riadku platí pre každé $i \in \{1, 2, \ldots, r\}$, môžeme písať

$$\sum_{i=1}^{r} 2^{-d_i} \leq \sum_{i=1}^{r} p_i \leq 1 \ .$$

Prirodzené čísla d_i pre $i = 1, 2, \ldots, r$ spĺňajú Kraftovu nerovnosť, a preto existuje binárne prefixové kódovanie s dĺžkami kódových slov $d_1, d_2, \ldots, d_r$. Pre strednú dĺžku slova tohoto kódovania platí:

$$d = \sum_{i=1}^{r} p_i.d_i < -\sum_{i=1}^{r} p_i.\big[\log_2(p_i) + 1\big] = -\sum_{i=1}^{r} p_i.\log_2(p_i) + \sum_{i=1}^{r} p_i = H(\mathcal{Z}) + 1 \ .$$

Nech d_{opt} je dĺžka najkratšieho prefixového binárneho kódovania abecedy A. Potom platí $H(\mathcal{Z}) \leq d_{\text{opt}} \leq d < H(\mathcal{Z}) + 1$.
Dokázané môžeme zhrnúť do nasledujúcej vety:

Veta 4.4. *Nech $\mathcal{Z}$ je stacionárny nezávislý zdroj s entropiou $H(\mathcal{Z})$, nech d_{opt} je stredná dĺžka kódového slova najkratšieho binárneho prefixového kódovania abecedy A. Potom platí:*

$$H(\mathcal{Z}) \leq d_{\text{opt}} < H(\mathcal{Z}) + 1 \ . \tag{4.16}$$

Príklad 4.4. Majme stacionárny nezávislý zdroj $\mathcal{Z}$ so zdrojovou abecedou $A = \{x, y, z\}$ s troma znakmi, ktorých pravdepodobnosti výskytu sú $p_x = 0.8$, $p_y = 0.1$, $p_z = 0.1$. Kódovanie $K(x) = 0$, $K(y) = 10$, $K(z) = 11$ je najkratšie binárne prefixové kódovanie abecedy A s dĺžkou $d(K) = 1{\times}0.8{+}2{\times}0.1{+}2{\times}0.1 = 1.2$. Entropia zdroja $\mathcal{Z}$ je $H(\mathcal{Z}) = 0.922$ bitu na znak. Ak máme dostatočne dlhý N-znakový zdrojový text, potom dĺžku príslušného zakódovaného textu možno odhadnúť číslom $N \times 1.2$, jej dolná hranica hranica určená podľa vety 4.4 je $N \times 0.922$. Dlhý zakódovaný zdrojový text bude teda v tomto prípade o 30% dlhší ako dolný odhad jeho dĺžky určený entropiou $H(\mathcal{Z})$.

Dal by sa nájsť ešte krikľavejší príklad percentuálnej odchýlky dolnej hranice určenej entropiou zdroja a dĺžkou optimálneho prefixového kódovania (skúste $p_x = 0.98$, $p_y = 0.01$, $p_z = 0.01$). Pretože žiadne jednoznačne dekódovateľné kódovanie zdrojovej abecedy A nemôže mať menšiu strednú dĺžku kódového slova, tento príklad nás nenapĺňa prílišným optimizmom čo sa týka užitočnosti dolného odhadu vo vete 4.4.

Kódovanie znak po znaku však nie jediným spôsobom, ako zakódovať zdrojový text. V časti 3.4 na str. 59 bol v definícii 3.8 k zdroju $\mathcal{Z}$ s entropiou $H(\mathcal{Z})$ definovaný zdroj $\mathcal{Z}_{(k)}$ s entropiou $k.H(\mathcal{Z})$, ktorý má za zdrojovú abecedu množinu všetkých k-znakových slov. V prípade, že $\mathcal{Z}$ je stacionárny nezávislý zdroj, je zdroj $\mathcal{Z}_{(k)}$ tiež stacionárnym nezávislým zdrojom. Pre strednú dĺžku $d_{\text{opt}}^{(k)}$ kódového slova najkratšieho binárneho prefixového kódovania abecedy A^k platí vzťah (4.16) z vety 4.4:

$$\begin{aligned} H(\mathcal{Z}_{(k)}) \leq & d_{\text{opt}}^{(k)} < H(\mathcal{Z}_{(k)}) + 1 \\ k.H(\mathcal{Z}) \leq & d_{\text{opt}}^{(k)} < k.H(\mathcal{Z}) + 1 \\ H(\mathcal{Z}) \leq & \frac{d_{\text{opt}}^{(k)}}{k} < H(\mathcal{Z}) + \frac{1}{k} \end{aligned} \tag{4.17}$$

Práve dokázané poznatky zhrnieme v nasledujúcej vete:

Veta 4.5. Základná veta o kódovaní zdrojov. *Nech $\mathcal{Z} = (A^*, P)$ je stacionárny nezávislý zdroj s entropiou $H(\mathcal{Z})$. Potom je stredná dĺžka zakódovaného binárneho textu pripadajúca na jeden znak zdrojovej abecedy A zdola ohraničená entropiou $H(\mathcal{Z})$. Pritom sa dá nájsť prirodzené číslo k a binárne prefixové kódovanie slov z A^k také, že stredná dĺžka zakódovaného textu pripadajúca na jeden znak zdrojovej abecedy A je ľubovoľne blízko entropii $H(\mathcal{Z})$.*

Základná veta o kódovaní zdrojov platí aj pre všeobecnejšie stacionárne zdroje $\mathcal{Z}$ (dôkaz je však o čosi zložitejší). Jej význam je v tom, že ukazuje entropiu zdroja ako limitnú hodnotu strednej dĺžky binárne zakódovaného textu

pripadajúceho na jeden znak zdrojovej abecedy. Ukazuje sa tu, že pojem entropie bol dobre zvolený a má svoj hlboký význam. Všimnime si tiež, že pre binárne kódovanie vo vzťahu (4.16) vo vete 4.4 vystupuje entropia $H(\mathcal{Z})$ bez prepočítavacieho koeficientu (resp. s koeficientom 1), čo je dôsledok toho, že sme šťastne zvolili číslo 2 za základ logaritmu pri Shannonovej definícii informácie, resp. pri Shannonovej – Hartleyovej formule pre entropiu.

Ako sme už ukázali, prirodzený jazyk ani zďaleka nemožno považovať za nezávislý zdroj, jeho entropia je oveľa menšia, ako entropia prvého písmena $H_1 = -\sum_i p_i \log_2(p_i)$. V takýchto prípadoch by sme mohli dostať kratšiu dĺžku zakódovanej správy tak, že by sme za znaky zdrojovej abecedy brali dvojice, trojice, prípadne n-tice pôvodnej abecedy A. Huffmanove kódovanie je základom mnohých metód kompresie údajov, kde sa k správe zakódovanej v blokovom binárnom kóde hľadá efektívnejší spôsob uloženia.

4.7 Kódy objavujúce chyby

V tejto časti sa budeme zaoberať blokovými kódmi s n-prvkovou kódovou abecedou a mnohokrát špeciálne dekadickými číslami. Do spracovanie takýchto prirodzených kódov býva včlenený aj ľudský činiteľ, ktorý je zdrojom častých chýb. Otázkou teraz je, či je možné nájsť kód taký, ktorý by zistil, že pri prenose nastala jedna, alebo aj viac chýb istého druhu. Z anglosaskej literatúry máme údaje o relatívnej početnosti chýb vznikajúcich písaním textov na klávesnici resp. písacom stroji.

- Jednoduchá chyba $a \to b$ 79%
- Susedná transpozícia $ab \to ba$ 10.2%
- Skoková transpozícia $abc \to cba$ 0.8%
- Blíženci $aa \to bb$ 0.6%
- Fonetická chyba $X0 \to 1X$ 0.5%
- Ostatné chyby 8.9%

Vidíme, že najbežnejšie ľudské chyby sú jednoduchá chyba a zámena poradia dvoch susedných písmen. Medzi ostatnými chybami môže byť aj vynechanie či pridanie znaku, avšak blokový kód takúto chybu ihneď odhalí, lebo mení dĺžku kódového slova. Fonetická chyba je pravdepodobne anglickou špecialitou

a vychádza z malého rozdielu medzi anglickými číslovkami (napr. fourteen – fourty, fifteen – fifty a pod.).

Ak má kódová abeceda B n znakov, potom počet všetkých slov dĺžky k je n^k – to je najväčší možný počet slov blokového kódu dĺžky k s n kódovými znakmi. Jediným spôsobom, ako na strane prijímača zistiť chybu v prijatom slove je tento: Pre **kódové slová** využiť len časť z n^k možných slov, ostatné slová prehlásiť za **nekódové**. Ak prijmeme nekódové slovo, vieme, že sme prijali slovo s chybou. Problémom však je, ako určiť množinu kódových slov tak, aby pri jednej alebo i viacerých chybách istého druhu vzniklo z kódového slova nekódové slovo, a ako rýchlo určiť, či prijaté slovo je alebo nie je kódové.

Pri študovaní tejto problematiky sa najskôr obmedzíme na jednoduché chyby, ktorým niekedy hovorím preklepy. Pre štúdium takýchto chýb je veľmi užitočné zaviesť si na množine $B^n \times B^n$ – dvojíc n-znakových slov funkciu, ktorá by vyjadrovala odlišnosť dvojice ľubovoľných takýchto slov. Najradšej by sme boli, keby táto funkcia mala vlastnosti analogické vlastnostiam vzdialenosti bodov v rovine či priestore.

Definícia 4.5. Reálna funkcia d definovaná na karteziánskom súčine $V \times V$ sa nazýva **metrikou na množine** V, ak platí:

1. Pre každé $u, v \in V$ je $d(u, v) \geq 0$ a rovnosť nastáva práve vtedy, keď $u = v$.
2. Pre každé $u, v \in V$ je $d(u, v) = d(v, u)$.
3. Ak $u, v, w \in V$, potom $d(u, w) \leq d(u, v) + d(v, w)$.

Definícia 4.6. Hammingova vzdialenosť $d(\mathbf{v}, \mathbf{w})$ dvoch slov $\mathbf{v} = v_1 v_2 \dots v_n$, $\mathbf{w} = w_1 w_1 \dots w_n$ je počet miest, na ktorých sa znaky slov $\mathbf{v}$, $\mathbf{w}$ líšia, t. j.

$$d(\mathbf{v}, \mathbf{w}) = \big|\{i \mid v_i \neq w_i, \quad i = 1, 2, \dots, n\}\big|.$$

Ľahko sa dá ukázať, že Hammingova vzdialenosť má vlastnosti metriky (čo si čitateľ môže urobiť sám), a preto sa niekedy volá aj **Hammingova metrika**.

Definícia 4.7. Minimálna vzdialenosť $\Delta(\mathcal{K})$ **blokového kódu** (K) je minimum zo vzdialeností všetkých dvojíc rôznych slov kódu $\mathcal{K}$.

$$\Delta(\mathcal{K}) = \min\{d(\mathbf{a}, \mathbf{b}) \mid \mathbf{a}, \mathbf{b} \in \mathcal{K}, \ \mathbf{a} \neq \mathbf{b}\}. \tag{4.18}$$

Hovoríme, že kód $\mathcal{K}$ **objavuje** t**-násobné jednoduché chyby**, ak pri zmene ľubovoľných t znakov kódového slova $\mathbf{u}$ vznikne nekódové slovo.

Ak teda prijmeme nekódové slovo, hovoríme, že sme objavili chybu. Všimnime si, že blokový kód $\mathcal{K}$ s minimálnou vzdialenosťou $\Delta(\mathcal{K}) = d$ objavuje $(d-1)$-násobné jednoduché chyby.

Príklad 4.5 (Kód dva z piatich)**.** Dva prvky z piatich možno vybrať $\binom{5}{2} = 10$ spôsobmi, čo možno využiť pre kódovanie dekadických cifier nasledovne:

1	11000	6	00101
2	10100	7	00011
3	10010	8	00110
4	10001	9	01100
5	01001	0	01010

Kód dva z piatich objavuje jednu chybu – pri zmene ktorejkoľvek 0 na 1 vznikne nekódové slovo s troma znakmi 1, pri zmene 1 na 0 dostaneme nekódové slovo obsahujúce len jeden znak 1. Kódové slová 11000 a 10100 majú však Hammingovu vzdialenosť rovnajúcu sa 2, z čoho vyplýva, že kód dva z piatich neobjavuje všetky 2-násobné jednoduché chyby.

Príklad 4.6. Kód s kontrolou parity je osembitový kód, kde prvých 7 bitov je ľubovoľný 7-miestny binárny blokový kód a kde je posledný binárny znak doplnený tak, aby počet jednotkových bitov bol párny. Kód s kontrolou parity objavuje jednu jednoduchú chybu, jeho minimálna vzdialenosť je 2. Princíp kontroly paritou bol veľmi často používaný pri prenosoch a občas sa s ním stretneme aj v súčasnosti.

Príklad 4.7. Zdvojovací kód. Ide o kód párnej dĺžky, v ktorom sa každý znak opakuje dvakrát. Zdvojovací binárny kód dĺžky 6 má osem kódových slov:

000000 000011 001100 001111 110000 110011 111100 111111

Zdvojovací kód má minimálnu vzdialenosť 2, objavuje jednu jednoduchú chybu.

Príklad 4.8. Opakovací kód. Princípom opakovacieho kódu je niekoľkonásobné opakovanie toho istého znaku. Kódové slová sú len slová pozostávajúce z toho istého znaku – napr. 11111, 22222, ..., 99999, 00000. Opakovací kód K dĺžky n má minimálnu vzdialenosť $\Delta K = n$ a preto objavuje $(n-1)$-násobné jednoduché chyby. Všimnime si, že za predpokladu, že nastali maximálne dve chyby, pri opakovacom kóde dĺžky 5 vieme zrekonštruovať pôvodné slovo. Ak prijmeme 10191, za predpokladu vzniku maximálne dvoch chýb vieme, že bolo vyslané slovo 11111.

Príklad 4.9. Medzinárodné číslo vagónu je 12-miestne dekadické číslo tvaru

$$X\ X\ XX\ X\ XXX\ XXX\ X$$

Prvá cifra predstavuje číslo medzinárodného spoločenstva, druhá triedu vyhovovania medzinárodným predpisom, v tretej a štvrtej cifre je zakódovaný vlastník, piata cifra obsahuje kód základného triedenia vozňa, ďalšie trojčíslie predstavuje kód technickej špecifikácie vozňa, trojčíslie od deviatej po jedenástu cifru obsahuje poradové číslo vozňa a posledná dvanásta cifra je kontrolná číslica.
Majme číslo vagóna

$$a_1a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11}a_{12}$$

Kontrolná číslica a_{12} sa určí tak, aby ciferný súčet čísel

$$2a_1\ a_2\ 2a_3\ a_4\ 2a_5\ a_6\ 2a_7\ a_8\ 2a_9\ a_{10}\ 2a_{11}\ a_{12}$$

bol deliteľný číslom 10. Cifry na párnych a nepárnych miestach sa spracovávajú odlišne – evidentne tu vidieť snahu poistiť sa i proti susednej zámene. Nech na dvoch susedných miestach sú cifry C, D, nech C je na nepárnom mieste. Označme $\delta(Y)$ ciferný súčet čísla $2Y$ pre $Y = 0, 1, \ldots, 9$. Potom

$$\delta(Y) = \begin{cases} 2Y & \text{ak } Y \leq 4 \\ 2Y - 9 & \text{ak } Y > 4 \end{cases}$$

Hľadajme, pre ktoré hodnoty cifier C, D sa kontrolná číslica po ich susednej zámene nezmení. Aby sa kontrolná číslica pri susednej zámene cifier nezmenila, musí dať súčet $\delta(C)+D$ ten istý zvyšok pri delení desiatimi ako $\delta(D)+C$ a teda ich rozdiel musí byť deliteľný desiatimi.

$$\delta(C)+D-\delta(D)-C = \begin{cases} 2C + D - 2D - C = C - D & \text{ak } C \leq 4 \text{ a } D \leq 4 \\ 2C - 9 + D - 2D - C = C - D - 9 & \text{ak } C \geq 5 \text{ a } D \leq 4 \\ 2C + D - 2D + 9 - C = C - D + 9 & \text{ak } C \leq 4 \text{ a } D \geq 5 \\ 2C + 9 + D - 2D - 9 - C = C - D & \text{ak } C \geq 5 \text{ a } D \geq 5 \end{cases}$$

V prvom a štvrtom prípade rozdiel $C - D$ je deliteľný desiatimi práve vtedy, keď $C = D$, z čoho vyplýva, že kód rozpozná každú susednú zámenu takých cifier, že sú obe menšie než 5 alebo obe väčšie než 4.
V druhom prípade ak $C \geq 5$ a súčasne $D \leq 4$, potom $1 \leq (C - D) \leq 9$. Výraz $\delta(C) + D - \delta(D) - C$ sa v tomto prípade rovná $(C - D) - 9$. Posledný výraz môže byť deliteľný desiatimi len vtedy, keď $C - D = 9$, čo môže nastať len pre

dvojicu $C = 9$, $D = 0$.
V treťom prípade ak $C \leq 4$ a $D \geq 5$, $0 - 9 = -9 \leq (C - D) \leq 4 - 5 = -1$. Výraz $\delta(C) + D - \delta(D) - C$ sa v tomto prípade rovná $(C - D) + 9$. Posledný výraz môže byť deliteľný desiatimi len vtedy, keď $C - D = -9$, čo môže nastať len pre dvojicu $C = 0$, $D = 9$. Vidíme, že rovnica

$$\delta(C) + D - \delta(D) - C \equiv 0 \mod 10$$

má len dve riešenia, a to $C = 0$, $D = 9$ a $C = 9$, $D = 0$.

Kódovanie vozňov teda objaví jednu jednoduchú chybu alebo jednu susednú zámenu dvojice znakov rôznej od 09 a 90. Ak sa v čísle vozňa kdekoľvek zamení dvojica 09 za 90, resp. 90 za 09, kontrolná číslica sa nezmení a teda kód takúto chybu nezistí. Objavovanie ktorejkoľvek susednej zámeny sa však konštruktérom tohoto kódu nepodarilo zaistiť.

4.8 Elementárne metódy objavovania chýb

V tejto a nasledujúcej časti 4.9 sa budeme zoberať metódami objavovania chýb v prirodzených dekadických blokových kódoch dĺžky n. Kódová abeceda týchto kódov je množina $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$. Princíp týchto metód bude spočívať v tom, že kódové slová $\mathbf{a} = a_1 a_2 \ldots a_{n-1} a_n$ budú mať prvých $n - 1$ znakov ľubovoľných (sú určené na to, aby niesli informáciu), posledný znak a_n bude tzv. kontrolný znak. Slovo $\mathbf{a}$ bude kódovým slovom vtedy a len vtedy, keď jeho posledný znak a_n bude vyhovovať tzv. kontrolnej rovnici typu

$$f(a_1, a_2, \ldots, a_n) = c \,, \tag{4.19}$$

kde f je vhodná funkcia. Budeme hľadať takú funkciu f, aby platilo:
Ak slovo $\mathbf{a}' = a'_1 a'_2 \ldots a'_{n-1} a'_n$ vzniklo zo slova $\mathbf{a} = a_1 a_2 \ldots a_{n-1} a_n$ jednoduchým preklepom alebo zámenou susedných znakov, potom $f(\mathbf{a}) \neq f(\mathbf{a}')$.

4.8.1 Kódy s kontrolnou rovnicou mod 10

Pre dekadické kódy určíme kontrolnú číslicu a_n z rovnice

$$a_n \equiv -\sum_{i=1}^{n-1} w_i . a_i \mod 10 \,,$$

kde w_i sú vopred zvolené pevné čísla, $0 \leq w_i \leq 9$. Tento prístup možno ešte trochu zovšeobecniť tak, že kódové slová tvorené znakmi a_1 až a_n sú práve tie

slová, ktoré vyhovujú tzv. kontrolnej rovnici

$$\sum_{i=1}^{n} w_i.a_i \equiv c \mod 10 . \tag{4.20}$$

Ak sa v slove $a_1a_2 \ldots a_n$ zmení a_j na a'_j, bude sa ľavá strana kontrolnej rovnice 4.20 pre takto zmenené slovo rovnať

$$\sum_{i=1}^{n} w_i.a_i + w_j.a'_j - w_j.a_j \equiv c + w_j.(a'_j - a_j) \mod 10 .$$

Pravá strana rovnice 4.20 sa nezmení a príslušný kód neobjaví jednoduchú chybu, ak

$$w_j.(a'_j - a_j) \equiv 0 \mod 10 .$$

Posledná rovnica má jediné riešenie $a'_j = a_j$ práve vtedy, keď w_j nie je súdeliteľné s číslom 10. Na miestach w_i môžu byť len čísla 1, 3, 7 a 9.

Skúsme zistiť, či kód s kontrolnou rovnicou (4.20) objavuje susedné zámeny. Kód nezistí susednú zámenu znakov $x,\ y$ na miestach $i,\ i+1$ práve vtedy, keď

$$\begin{aligned} w_i.y + w_{i+1}.x - w_i.x - w_{i+1}.y &\equiv 0 \mod 10 \\ w_i.(y-x) - w_{i+1}.(y-x) &\equiv 0 \mod 10 \\ (w_i - w_{i+1})(y-x) &\equiv 0 \mod 10 \end{aligned}$$

K tomu, aby posledná rovnica nemala okrem riešení $x = y$ žiadne ďalšie je nutné a stačí, aby $(w_i - w_{i+1})$ bolo nesúdeliteľné s 10. Ak má však kód s kontrolnou rovnicou (4.20) rozoznávať jednoduché chyby, musí byť $w_i \in \{1, 3, 7, 9\}$ a preto je $(w_i - w_{i+1})$ vždy párne.

Veta 4.6. *Nech K je desiatkový blokový kód dĺžky n s kontrolnou rovnicou* (4.20)*. Kód K objavuje jednoduché chyby práve vtedy, keď sú všetky w_i nesúdeliteľné s 10, t. j. $w_i \in \{1, 3, 7, 9\}$. Žiaden desiatkový blokový kód dĺžky n s kontrolnou rovnicou* (4.20) *neobjavuje jednoduché chyby a súčasne aj susedné zámeny.*

Príklad 4.10. EAN European Article Number je 13-miestny dekadický kód, ktorým sa jedinečne označujú výrobky v rámci Európy. EAN kód býva prevedený do čiarového kódu, ktorý je umiestnený na obale výrobku. Pri manipulácii

s výrobkami sa tento kód sníma opticky, čím sa znižuje prácnosť pri evidencii, fakturácii, inventarizácii a ďalších operáciách s výrobkom.

Prvých dvanásť znakov a_1 až a_{12} kódu EAN je významových, znak a_{13} je kontrolný a vypočíta sa z rovnice

$$a_{13} \equiv -(1.a_1 + 3.a_2 + 1.a_3 + 3.a_4 + \cdots + 1.a_{11} + 3.a_{12}) \mod 10 \ .$$

Kód EAN odhaľuje jednoduché chyby. Pre dvojicu znakov x, y, na dvoch susedných miestach nepárnom a párnom kód neodhalí susednú zámenu, ak

$$\begin{aligned} (x + 3y) - (3x + y) &\equiv 0 \mod 10 \\ (2y - 2x) &\equiv 0 \mod 10 \\ 2.(y - x) &\equiv 0 \mod 10 \end{aligned}$$

Posledná rovnica má tieto riešenia (x, y):

$$\begin{aligned} &(0,0),\ (0,5),\ (1,1),\ (1,6),\ (2,2),\ (2,7),\ (3,3),\ (3,8),\ (4,4),\ (4,9), \\ &(5,5),\ (5,0),\ (6,6),\ (6,1),\ (7,7),\ (7,2),\ (8,8),\ (8,3),\ (9,9),\ (9,4) \end{aligned}$$

Pre desať usporiadaných dvojíc rôznych znakov kód EAN neobjaví susednú zámenu. Z hľadiska počtu neobjavených chýb je na tom horšie ako medzinárodný číselník vozňov, pre ktorý je neobjaviteľná iba susedná zámena znakov $(0, 9)$ a $(9, 0)$.

4.8.2 Kontrola modulo 11

Tieto kódy pracujú s kódovou abecedou $B \cup \{X\}$, kde znak X nahradzuje číslicu 10, $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$, pričom kódové slová majú všetkých prvých $n - 1$ znakov z abecedy B a posledný – kontrolný znak a_n z abecedy $B \cup \{X\}$ je určený tak, aby platila nasledujúca kontrolná rovnica

$$\sum_{i=1}^{n} w_i.a_i \equiv c \mod 11 \ , \quad \text{kde } 0 < w_i \leq 10 \text{ pre } i = 1, 2, \ldots, n \ . \tag{4.21}$$

Keďže Podobne ako v prípade kontroly modulo 10 ukážeme, že kód kontroly modulo 11 objavuje jednoduché chyby na mieste j práve vtedy, keď rovnica

$$w_j.(a'_j - a_j) \equiv 0 \mod 11$$

nemá okrem $a'_j = a_j$ žiadne iné riešenia, a to je práve vtedy, keď w_j je nesúdeliteľné s 11 na čo stačí, aby $w_j \neq 0$.

Na to, aby kód s kontrolou modulo 11 objavoval susedné zámeny na miestach i, $i+1$ stačí, aby rovnica

$$(w_i - w_{i+1}).(y - x) \equiv 0 \mod 11$$

okrem riešení, kde $x = y$ nemala žiadne iné riešenia. Na to však stačí, aby $w_i \neq w_{i+1}$. Príkladom tohoto kódu je kód ISBN.
Nakoniec poznamenajme, že vlastnosť objavovania jednoduchých chýb a susedných zámen sa nestratí, ak dovolíme, aby všetky znaky kódových slov (nielen kontrolný znak) boli z abecedy $B \cup \{X\}$.

Príklad 4.11. ISBN – International Standard Book Number je 10 miestne číslo prideľované každej oficiálne vydanej knihe, v ktorom prvé štyri znaky $a_1a_2a_3a_4$ určujú krajinu a vydavateľstvo, ďalších päť znakov $a_5a_6a_7a_8a_9$ predstavuje číslo knihy v rámci špecifikovaného vydavateľstva a posledný znak a_{10} je kontrolný znak určený rovnicou

$$a_{10} \equiv \sum_{i=1}^{9} i.a_i \mod 11 \ .$$

Znaky a_1 až a_9 sú z abecedy $A = \{0, 1, \ldots, 9\}$, znak a_{10} je z abecedy $A \cup \{X\}$, kde znak X predstavuje hodnotu 10.
Posledná rovnica je totožná s rovnicou

$$\sum_{i=1}^{10} i.a_i \equiv 0 \mod 11 \ ,$$

pretože $-a_{10} \equiv -a_{10} + 11.a_{10} \equiv 10.a_{10} \mod 11$. Ak $a_{10} = 10$, píše sa na mieste a_{10} znak X. Toto je istá nevýhoda ISBN kódovania, pretože kódová abeceda je 11-prvková, ale znak X sa využíva len zriedka. ISBN kód objavuje všetky jednoduché chyby a všetky susedné zámeny.

Mnoho dobrých vlastností má tzv. **geometrický kód modulo 11**, kde čísla w_i v kontrolnej rovnici (4.21) sú určené ako

$$w_i = 2^i \mod 11 \ .$$

Príklad 4.12. Čísla bankových účtov slovenských bánk. Číslo bankového účtu je desaťmiestne dekadické číslo

$$a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9.$$

Význam jednotlivých pozícií Národná banka Slovenska nešpecifikuje, avšak pre všetky banky platí rovnaký kontrolný mechanizmus. Platné číslo bankového účtu musí vyhovovať kontrolnej rovnici

$$0 = \left(\sum_{i=0}^{9} 2^i.a_i\right) \mod 11 = (1.a_0+2.a_1+4.a_2+8.a_3+\cdots+512.a_9) \mod 11 =$$
$$= (a_0 + 2a_1 + 4a_2 + 8a_3 + 5a_4 + 10a_5 + 9a_6 + 7a_7 + 3a_8 + 6a_9) \mod 11 .$$

Tu je použitý geometrický kód modulo 11.
Kód bankových účtov teda odhaľuje okrem jednoduchých chýb aj všetky susedné zámeny, ba navyše aj vzájomné zámeny znakov na rôznych pozíciách čísla účtu.

Príklad 4.13. Rodné číslo. Na stránke `www.minv.sk/vediet/rc.html` je uvedená nasledujúca špecifikácia:
Rodné číslo je definované v zákone ako číselný identifikačný osobný údaj, vytvorený z dátumu narodenia osoby, jej pohlavia a rozlišovacej koncovky.

Rodné číslo má tvar RRMMDDKKKK, kde

- RR vyjadruje posledné dve číslice roku narodenia osoby.
- MM vyjadruje mesiac narodenia a pohlavie osoby (napr. pre muža narodeného v januári je to dvojčíslie 01 pre ženu 51, pre muža narodeného v decembri je to dvojčíslie 12 a pre ženu 62).
- DD vyjadruje deň narodenia osoby (napr. pre osoby narodené v 1. deň mesiaca je to dvojčíslie 01, pre osoby narodené 15. deň je to dvojčíslie 15).
- KKKK je rozlišujúca koncovka pre osoby narodené v ten istý deň. Pre osoby narodené pred 1.1.1954 je koncovka trojmiestne číslo, pre osoby narodené po 31.12.1954 je koncovka štvormiestne číslo, čiže osoba narodená pred 1.1.1954 má deväťmiestne rodné číslo a osoba narodená po 31.12.1953 má desaťmiestne rodné číslo. Napríklad muž narodený 31.12.1925 môže mať rodné číslo 251231 123, žena narodená v ten istý deň 256231 123, muž narodený 1.1.1954 môže mať rodné číslo 540101 4311, žena 545101 1324. Desaťmiestne rodné číslo musí byť deliteľné číslom 11, pre deväťmiestne rodné číslo uvedená podmienka neplatí.

Majme desaťmiestne rodné číslo $a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9$. Skúmajme, aké druhy chýb je možno u rodných čísel objaviť.

Z podmienky deliteľnosti rodného čísla číslom 11 vyplýva nasledujúca kontrolná rovnica:

$$\sum_{i=0}^{9} 10^i.a_i \equiv 0 \mod 11 \ .$$

Ak je i párne číslo, t. j. $i = 2k$, potom $10^i = 10^{2k} = 100^k = (99+1)^k$. Podľa binomickej vety môžeme písať

$$(99+1)^k = \binom{k}{k}99^k + \binom{k}{k-1}99^{k-1} + \cdots + \binom{k}{1}99^1 + 1 \ . \tag{4.22}$$

Keďže číslo 99 je deliteľné jedenástimi, z posledného vyjadrenia máme

$$10^i \equiv 1 \mod 11 \quad \text{pre } i \text{ párne.}$$

Ak je i nepárne číslo, t. j. $i = 2k+1$, potom $10^i = 10^{2k+1} = 10.100^k = 10.(99+1)^k$. S využitím (4.22) môžeme písať

$$\begin{aligned} 10.(99+1)^k &= 10.\left[\binom{k}{k}99^k + \binom{k}{k-1}99^{k-1} + \cdots + \binom{k}{1}99^1 + 1\right] = \\ &= 10.\binom{k}{k}99^k + 10.\binom{k}{k-1}99^{k-1} + \cdots + 10.\binom{k}{1}99^1 + 10 \ . \end{aligned}$$

Z posledného vzťahu máme

$$10^i \equiv 10 \mod 11 \quad \text{pre } i \text{ nepárne.}$$

Kontrolná rovnica desaťmiestneho rodného čísla má teda tvar

$$a_0 + 10a_1 + a_2 + 10a_3 + a_4 + 10a_5 + a_6 + 10a_7 + a_8 + 10a_9 \equiv 0 \mod 11 \ , \tag{4.23}$$

odkiaľ už na základe doterajších poznatkov o kontrole modulo 11 môžeme tvrdiť, že kód desaťmiestnych rodných čísel objavuje okrem jednoduchých chýb aj susedné zámeny[1].

[1] Čitateľ si ľahko overí, že ekvivalentná rovnica s kontrolnou rovnicou (4.23) je rovnica

$$a_0 - a_1 + a_2 - a_3 + a_4 - a_5 + a_6 - a_7 + a_8 - a_9 \equiv 0 \mod 11.$$

4.9 Kódovanie s kontrolným znakom nad grupou*

Pri kódoch s kontrolnou rovnicou modulo 10 kódy objavovali jednoduché chyby práve vtedy, keď zobrazenie $a_i \mapsto \delta(a_i) = (w_i.a_i \mod 10)$ (zvyšok po delení čísla $w_i.a_i$ desiatimi) bolo permutáciou. Dokonca aj priradenie $x \mapsto \delta(x) =$ ciferný_súčet$(2.x)$ v kódovaní železničných vozňov je permutáciou (v príklade 4.14 je to permutácia δ_1) a tam sme dosiahli doteraz najlepší výsledok, čo sa týka zabezpečenia proti susedným zámenám. Vzniká teda myšlienka členy $w_i.a_i$ kontrolnej rovnice nahradiť permutáciami $\delta_i(a_i)$.

Príklad 4.14. Medzinárodné číslo vozňa je vlastne kód s permutáciami

$$\delta_1 = \delta_3 = \cdots = \delta_{11} := \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 0\,2\,4\,6\,8\,1\,3\,5\,7\,9 \end{pmatrix}$$
$$\delta_2 = \delta_4 = \cdots = \delta_{12} := \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \end{pmatrix}$$

a s kontrolnou rovnicou

$$\sum_{i=1}^{12} \delta_i(a_i) \equiv 0 \mod 10 \,.$$

Príklad 4.15. Kód nemeckých poštových poukážok je desaťmiestny dekadický kód $a_1a_2 \ldots a_{10}$ s kontrolným znakom a_{10} s kontrolnou rovnicou

$$\sum_{i=1}^{10} \delta_i(a_i) \equiv 0 \mod 10 \,,$$

kde

$$\delta_1 = \delta_4 = \delta_7 = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 1\,2\,3\,4\,5\,6\,7\,8\,9\,0 \end{pmatrix} \qquad \delta_2 = \delta_5 = \delta_8 = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 2\,4\,6\,8\,0\,1\,3\,5\,7\,9 \end{pmatrix}$$
$$\delta_3 = \delta_6 = \delta_9 = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 3\,6\,9\,1\,4\,7\,0\,2\,5\,8 \end{pmatrix} \qquad \delta_{10} = \begin{pmatrix} 0\,1\,2\,3\,4\,5\,6\,7\,8\,9 \\ 0\,9\,8\,7\,6\,5\,4\,3\,2\,1 \end{pmatrix}$$

Ani jeden z uvedených kódov kontroly modulo 10 neobjavuje všetky susedné zámeny. Preto ďalším zovšeobecnením je nahradenie grupy zvyškových tried

s grupovou operáciou $a \oplus b = (a + b) \mod 10$ nejakou inou grupou $\mathbb{G} = (A, *)$ a kontrolnú rovnicu formulovať ako

$$\prod_{i=1}^{n} \delta_i(a_i) = c \ . \tag{4.24}$$

Multiplikatívny tvar grupovej operácie $*$ naznačuje, že grupa $\mathbb{G}$ nemusí byť komutatívna.

Definícia 4.8. Nech A je abeceda, nech $\mathbb{G} = (A, *)$ je grupa. Nech $\delta_1, \delta_2, \ldots, \delta_n$, sú permutácie na A. Potom kontrolnou rovnicou (4.24) definovaný kód nazveme **kód s kontrolným znakom nad grupou** $\mathbb{G}$.

Permutácie sú vzájomne jednoznačné zobrazenia A na A. Preto ku každej permutácii δ existuje inverzná permutácia označovaná δ^{-1} pre ktorú platí

$$\delta(a) = x \quad \text{práve vtedy, keď} \quad \delta^{-1}(x) = a \ .$$

Ak máme dve permutácie δ_i, δ_j potom predpisom $\forall a \in A \ a \mapsto \delta_i\big(\delta_j(a)\big)$ vznikne permutácia, ktorú budeme značiť $\delta_i \circ \delta_j$, teda

$$\delta_i \circ \delta_j(a) = \delta_i\big(\delta_j(a)\big) \quad \forall a \in A \ .$$

Veta 4.7. *Aby kód $\mathcal{K}$ s kontrolným znakom nad grupou $\mathbb{G} = (A, *)$ rozpoznal zámenu ľubovoľných susedných znakov na miestach i, $i+1$ je nevyhnutné a stačí, aby*

$$x * \delta_{i+1} \circ \delta_i^{-1}(y) \neq y * \delta_{i+1} \circ \delta_i^{-1}(x) \tag{4.25}$$

pre všetky $x \in A$, $y \in A$, $x \neq y$.

Pre Abelovu grupu $\mathbb{G} = (A, +)$ možno vzťah (4.25) prepísať v tvare $x + \delta_{i+1} \circ \delta_i^{-1}(y) \neq y + \delta_{i+1} \circ \delta_i^{-1}(x)$, odkiaľ máme nasledujúci dôsledok:

Dôsledok. *Kód $\mathcal{K}$ s kontrolným znakom nad Abelovou grupou $\mathbb{G} = (A, +)$ objavuje zámenu ľubovoľných susedných znakov na miestach i, $i+1$ práve vtedy, keď pre ľubovoľné $x, \ y \in A$, $x \neq y$ platí:*

$$x - \delta_{i+1} \circ \delta_i^{-1}(x) \neq y - \delta_{i+1} \circ \delta_i^{-1}(y). \tag{4.26}$$

Dôkaz. Nech kód $\mathcal{K}$ rozpoznáva susednú zámenu na miestach i, $i + 1$. Potom pre ľubovoľné a_i, a_{i+1} také, že $a_i \neq a_{i+1}$ platí:

$$\delta_i(a_i) * \delta_{i+1}(a_{i+1}) \neq \delta_i(a_{i+1}) * \delta_{i+1}(a_i) \tag{4.27}$$

Pre ľubovoľné $x \in A$ existuje nejaké $a_i \in A$ také, že , potom $a_i = \delta_i^{-1}(x)$. Podobne pre ľubovoľné $y \in A$ existuje nejaké $a_{i+1} \in A$ také, že , potom $a_{i+1} = \delta_i^{-1}(y)$. Dosaďme do (4.27) najprv x za $\delta_i(a_i)$ a y za $\delta_i(a_{i+1})$, potom $\delta_i^{-1}(x)$ za a_i a $\delta_i^{-1}(y)$ za a_{i+1}. Dostaneme

$$\begin{aligned} x * \delta_{i+1}(a_{i+1}) &\neq y * \delta_{i+1}(a_i) \\ x * \delta_{i+1}\big(\delta_i^{-1}(y)\big) &\neq y * \delta_{i+1}\big(\delta_i^{-1}(x)\big) \\ x * \delta_{i+1} \circ \delta_i^{-1}(y) &\neq y * \delta_{i+1} \circ \delta_i^{-1}(x) \end{aligned}$$

a teda platí (4.25).

Nech platí (4.25) pre všetky $x,\ y \in A$, $x \neq y$. Potom (4.25) platí aj pre $x = \delta_i(a_i)$, $y = \delta_i(a_{i+1})$, kde $a_i,\ a_{i+1} \in A$, $a_i \neq a_{i+1}$.

$$\begin{aligned} \delta_i(a_i) * \delta_{i+1} \circ \delta_i^{-1}(\delta_i(a_{i+1})) &\neq \delta_i(a_{i+1}) * \delta_{i+1} \circ \delta_i^{-1}(\delta_i(a_i)) \\ \delta_i(a_i) * \delta_{i+1}\Big(\underbrace{\delta_i^{-1}\big(\delta_i(a_{i+1})\big)}_{a_{i+1}}\Big) &\neq \delta_i(a_{i+1}) * \delta_{i+1}\Big(\underbrace{\delta_i^{-1}\big(\delta_i(a_i)\big)}_{a_i}\Big) \\ \delta_i(a_i) * \delta_{i+1}(a_{i+1}) &\neq \delta_i(a_{i+1}) * \delta_{i+1}(a_i)\ , \end{aligned}$$

z čoho vyplýva, že kód $\mathcal{K}$ objavuje susednú zámenu na miestach i, $i+1$. ■

Všimnime si vzťah (4.26). Ten hovorí, že priradenie $x \mapsto \big(x - \delta_{i+1} \circ \delta_i^{-1}(x)\big)$ je prosté – je tiež permutáciou.

Definícia 4.9. Permutácia δ (multiplikatívnej) grupy $\mathbb{G} = (A, *)$ sa nazýva **úplným zobrazením**, ak zobrazenie definované vzťahom

$$\forall x \in A \quad x \mapsto \eta(x) = x * \delta(x)$$

je zase permutácia.
Permutácia δ (aditívnej) grupy $\mathbb{G} = (A, +)$ sa nazýva **úplným zobrazením**, ak ak zobrazenie definované vzťahom

$$\forall x \in A \quad x \mapsto \eta(x) = x + \delta(x)$$

je zase permutácia.

Veta 4.8. *Kód $\mathcal{K}$ s kontrolným znakom nad Abelovou grupou $\mathbb{G} = (A, +)$ objavujúci jednoduché chyby a susedné zámeny existuje práve vtedy, keď existuje úplné zobrazenie grupy $\mathbb{G}$.*

Dôkaz. Definujme zobrazenie $\mu : A \to A$ predpisom $\mu(x) = -x$. Zobrazenie μ je prosté - je to permutácia. Pre ľubovoľnú permutáciu δ množiny A je zobrazenie $x \mapsto -\delta(x) = \mu \circ \delta(x)$ zase permutáciou.

Nech kód $\mathcal{K}$ objavuje susedné zámeny, potom podľa dôsledku vety 4.7 je zobrazenie $x \mapsto \left(x - \delta_{i+1} \circ \delta_i^{-1}(x)\right)$ permutáciou. Ale

$$x - \delta_{i+1} \circ \delta_i^{-1}(x) = x + \underbrace{\mu \circ \delta_{i+1} \circ \delta_i^{-1}(x)}_{\delta(x)} = x + \delta(x)$$

Permutácia δ definovaná predpisom $\delta = \mu \circ \delta_{i+1} \circ \delta_i^{-1}$ je hľadaným úplným zobrazením.

Nech existuje úplné zobrazenie δ grupy $\mathbb{G}$. Definujme

$$\delta_i = (\mu \circ \delta)^i. \tag{4.28}$$

Potom

$$x - \delta_{i+1} \circ \delta_i^{-1}(x) = x - (\mu \circ \delta)^{i+1} \circ (\mu \circ \delta)^{-i}(x) = x - (\mu \circ \delta)(x) = x + \delta(x),$$

z čoho vyplýva, že $x - \delta_{i+1} \circ \delta_i^{-1}(x)$ je permutácia. Podľa dôsledku vety 4.7 kód s kontrolným znakom nad grupou $\mathbb{G}$ s permutáciami δ_i definovanými v (4.28) objavuje susedné zámeny. ■

Veta 4.9. *Nech $\mathbb{G}$ je Abelova konečná grupa. Potom platí nasledujúca veta (pozri [11], 8.11 str. 63):*

a) Ak je $\mathbb{G}$ grupa nepárneho rádu, potom je identita na $\mathbb{G}$ úplným zobrazením.

b) Grupa $\mathbb{G}$ rádu $r = 2.m$, kde m je nepárne číslo, nemá žiadne úplné zobrazenie

c) Nech $\mathbb{G} = (A, +)$ je Abelova grupa párneho rádu. Potom na $\mathbb{G}$ existuje úplné zobrazenie práve vtedy, keď grupa obsahuje aspoň dve rôzne involúcie, t. j. také prvky $g \in A$, že $g \neq 0$, a $g + g = 0$

Dôkaz. Dôkaz tejto vety patrí do teórie konečných grúp, presahuje zámery tejto publikácie, preto vetu uvádzame bez dôkazu.

Ak máme abecedu $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ na ktorej definujeme grupovú operáciu $\oplus$ predpisom $a \oplus b = (a + b) \mod 10$, možno všetky kódy s kontrolnou cifrou modulo 10 interpretovať ako kódy s kontrolným znakom nad Abelovou grupou $\mathbb{G} = (A, \oplus)$. Táto grupa je rádu $r = 10 = 2.5$, a preto v nej neexistuje úplné zobrazenie.

Dôsledok. Neexistuje žiaden dekadický kód s kontrolným znakom nad Abelovou grupou $\mathbb{G} = (A, \oplus)$, kde $A = \{0,1,2,3,4,5,6,7,8,9\}$, ktorý by objavoval jednoduché chyby a susedné zámeny.

Jediná šanca na zostrojenie dekadického kódu, ktorý by objavoval jednoduché chyby a susedné zámeny je skúsiť kód s kontrolným znakom nad nekomutatívnou grupou.

Definícia 4.10. Diederova grupa $\mathbb{D}_n$ je konečná grupa rádu $2.n$ tvaru

$$\{1, a, a^2, \ldots, a^{n-1}, b, ab, a^2.b, \ldots, a^{n-1}b\},$$

kde platí

$$\begin{aligned} a^n &= 1 \quad (a^i \neq 1 \text{ pre } i = 1,2,\ldots,n-1 \\ b^2 &= 1 \quad (b \neq 1) \\ b.a &= a^{n-1}.b \end{aligned}$$

Diederovu grupu $\mathbb{D}_n$ budeme značiť

$$\mathbb{D}_n = \langle a, b \mid a^n = 1 = b^2,\ ba = a^{n-1}b \rangle$$

Diederovu grupu $\mathbb{D}_n$ možno interpretovať ako grupu symetrií pravidelného n-uholníka – prvok a ako rotáciu okolo stredu o uhol $2\pi/n$, prvok b ako osovú súmernosť. Pre $\mathbb{D}_3$ $1 = (ABC)$, $a = (CAB)$, $a^2 = (BCA)$, $b = (ACB)$, $ab = (BAC)$, $a^2b = (CBA)$.

Príklad 4.16. Diederova grupa $\mathbb{D}_3 = \langle a, b \mid a^3 = 1 = b^2,\ ba = a^2b \rangle$. Prvky grupy $\mathbb{D}_3$ možno priradiť číslam od 1 po 6 aj takto:

1	a	a^2	b	ab	a^2b
1	2	3	4	5	6

Označme grupovú operáciu v $\mathbb{D}_3$ symbolom $\otimes$. Potom počítame

$$2 \otimes 3 = a.a^2 = a^3 = 1$$

$$3 \otimes 6 = a^2.a^2b = a^4b = a^3.ab = 1.ab = ab = 5$$

$$6 \otimes 3 = a^2b.a^2 = ba.a^2 = ba^3 = b.1 = b = 4$$

$$4 \otimes 5 = b.ab = ba.b = a^2b.b = a^2.b^2 = a^2.1 = a^2 = 3$$

$$5 \otimes 4 = ab.b = a.b^2 = a.1 = a = 2$$

Veta 4.10. *Nech* $\mathbb{D}_n = \langle a, b \mid a^n = 1 = b^2,\ ba = a^{n-1}b \rangle$ *je Diederova grupa nepárneho rádu* n, $n \geq 3$. *Definujme permutáciu* $\delta : \mathbb{D}_n \to \mathbb{D}_n$ *predpisom*

$$\delta(a^i) = a^{n-1-i} \quad a \quad \delta(a^i b) = a^i b \quad \forall i = 0, 1, 2, \ldots, n-1 \ . \tag{4.29}$$

Potom pre permutáciu δ *platí:*

$$x.\delta(y) \neq y.\delta(x) \quad \forall x, y \in \mathbb{D}_n \text{ také, že } x \neq y \ . \tag{4.30}$$

Dôkaz. Prv než sa pustíme do samotného dôkazu, uvedomme si jednu skutočnosť. Podľa definície Diederovej grupy je $b.a = a^{n-1}b$. Keďže $a^{n-1}.a = 1$, je $a^{n-1} = a^{-1}$, a preto platí $ba = a^{-1}b$. Nech k je ľubovoľné prirodzené číslo. Potom $b.a^k = a^{-1}ba^{k-1} = a^{-2}ba^{k-2} = \cdots = a^{-k}b$. Pre ľubovoľné celé číslo platí

$$b.a^k = a^{-k}b \ . \tag{4.31}$$

Ľahko sa overí, že δ je prosté zobrazenie – permutácia. Aby sme overili (4.30), budeme rozoznávať tri prípady.
1. prípad:
$x = a^i$, $y = a^j$, kde $i \neq j$, $0 \leq i, j \leq n-1$. Keby platilo $x.\delta(y) = y.\delta(x)$, potom by $a^i.a^{n-1-j} = a^j.a^{n-1-i}$, z čoho vyplýva $a^{2i-2j} = a^{2(i-j)} = 1$. Číslo $2(i-j)$ musí byť deliteľná nepárnym číslom n, keby totiž $2(i-j) = kn + r$, kde $1 \leq r \leq n-1$, potom by $a^{2(i-j)} = a^{kn+r} = a^{kn}a^r = 1.a^r \neq 1$. Ak má nepárne n deliť $2(i-j)$, musí byť $(i-j)$ deliteľné číslom n, čo môže nastať len tak, že $(i-j) = 0$, lebo $0 \leq i, j \leq n-1$.
2. prípad:
$x = a^i$, $y = a^j b$, $0 \leq i, j \leq n-1$. Nech $x.\delta(y) = y.\delta(x)$, t. j. $a^i a^j b = a^j b a^{n-1-i}$. Použitím (4.31) máme $a^{i+j}b = a^j.a^{i+1}b$ odkiaľ postupne dostaneme
$a^{i+j} = a^{i+j+1}$, $1 = a$. V definícii Diederovej grupy $\mathbb{D}_n$ pre $n \geq 3$ však $a \neq 1$.
3. prípad:
$x = a^i b$, $y = a^j b$, $0 \leq i, j \leq n-1$. Nech $x.\delta(y) = y.\delta(x)$, čo v tomto prípade znamená $a^i b.a^j b = a^j b a^i b$. S využitím (4.31) máme $a^i bb.a^{-j} = a^j bba^{-i}$. Pretože $b.b = b^2 = 1$ má posledná rovnica tvar $a^{i-j} = a^{j-i}$, čiže $a^{2(i-j)} = 1$. Pri riešení 1. prípadu sme však ukázali, že je to možné len ak $i = j$.

Veta 4.11. *Nech* $\mathbb{D}_n = \langle a, b \mid a^n = 1 = b^2,\ ba = a^{n-1}b \rangle$ *je Diederova grupa nepárneho rádu* n, $n \geq 3$. *Nech permutácia* $\delta : \mathbb{D}_n \to \mathbb{D}_n$ *je definovaná predpisom* (4.29). *Definujme permutácie* $\delta_i = \delta^i$ *pre* $i = 1, 2, \ldots, m$. *Potom blokový kód dĺžky* m *s kontrolným znakom nad grupou* $\mathbb{D}_n$ *objavuje jednoduché chyby a susedné zámeny.*

Dôkaz. Podľa vety 4.7 stačí ukázať, že pre $x \neq y$ platí

$$x * \delta_{i+1} \circ \delta_i^{-1}(y) \neq y * \delta_{i+1} \circ \delta_i^{-1}(x)$$

Keby pre nejaké $x \neq y$ v poslednom vzťahu platila rovnosť, dosadením za $\delta_i = \delta^i,\ \delta_{i+1} = \delta^{i+1}$ by sme dostali

$$x * \delta^{i+1} \circ \delta^{-i}(y) = y * \delta^{i+1} \circ \delta^i(x)$$
$$x * \delta(y) = y * \delta(x),$$

čo by bolo v spore s vlastnosťami (4.30) permutácie δ. ■

Poznámka. Definíciu (4.29) možno zovšeobecniť nasledovne: Definujme $\delta : \mathbb{D}_n \to \mathbb{D}_n$ predpisom

$$\delta(a^i) = a^{c-i+d} \quad \text{a} \quad \delta(a^i b) = a^{i-c+d} b \quad \forall i = 1, 2, \ldots, n-1 \tag{4.32}$$

Potom definícia (4.29) je špeciálnym prípadom (4.32) pre $c = d = \dfrac{n-1}{2}$.

Príklad 4.17. Diederova grupa $\mathbb{D}_5 = \langle a, b \mid a^5 = 1 = b^2,\ ba = a^4 b \rangle$. Prvky grupy $\mathbb{D}_5$ možno priradiť dekadickým znakom nasledovne:

1	a	a^2	a^3	a^4	b	ab	a^2b	a^3b	a^4b
0	1	2	3	4	5	6	7	8	9

Pre grupovú operáciu $i * j$ bude platiť nasledovná schéma

$i * j$	$0 \leq j \leq 4$	$5 \leq j \leq 9$
$0 \leq i \leq 4$	$(i + j) \mod 5$	$5 + [(i + j) \mod 5]$
$5 \leq i \leq 9$	$5 + [(i - j) \mod 5]$	$(i - j) \mod 5$

z ktorej dostaneme tabuľku pre operáciu $*$

		j									
	$*$	0	1	2	3	4	5	6	7	8	9
i	0	0	1	2	3	4	5	6	7	8	9
	1	1	2	3	4	0	6	7	8	9	5
	2	2	3	4	0	1	7	8	9	5	6
	3	3	4	0	1	2	8	9	5	6	7
	4	4	0	1	2	3	9	5	6	7	8
	5	5	9	8	7	6	0	4	3	2	1
	6	6	5	9	8	7	1	0	4	3	2
	7	7	6	5	9	8	2	1	0	4	3
	8	8	7	6	5	9	3	2	1	0	4
	9	9	8	7	6	5	4	3	2	1	0

4.10 Všeobecná teória kódov opravujúcich t jednoduchých chýb

Majme abecedu $A = \{a_1, a_2, \ldots, a_r\}$ s r znakmi. V tejto časti budeme skúmať blokové kódy dĺžky n, t. j. podmnožiny typu $\mathcal{K} \subset A^n$ z hľadiska všeobecných možnosti objavenia a opravy t jednoduchých chýb.

Podľa definície 4.6 sa Hammingova vzdialenosť $d(\mathbf{a}, \mathbf{b})$ dvoch slov $\mathbf{a}$, $\mathbf{b} \in A^n$ rovná počtu miest, na ktorých majú slová $\mathbf{a}$, $\mathbf{b}$ rôzne znaky. Minimálna vzdialenosť $\Delta\mathcal{K}$ sa podľa definície 4.7 (str. 83) rovná minimu zo vzdialeností všetkých dvojíc rôznych slov kódu $\mathcal{K}$. Kód $\mathcal{K}$ objavuje t-násobné jednoduché chyby, ak pri zmene ľubovoľných t znakov slova $\mathbf{c}$ vznikne nekódové slovo. Ak teda prijmeme nekódové slovo, hovoríme, že sme objavili chybu.

Maximum vzdialeností dvoch slov z A^n môže byť n – to v prípade, keď príslušné slová nemajú ani na jednom mieste rovnaký znak.

Veta 4.12. *Hammingova vzdialenosť je metrikou na A^n, t. j. platí:*

$$d(\mathbf{a}, \mathbf{b}) \geq 0 \; ; \quad d(\mathbf{a}, \mathbf{b}) = 0 \iff \mathbf{a} = \mathbf{b}$$
$$d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$$
$$d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$$

(A^n, d) je teda metrický priestor.

Dôkaz. Overenie vlastností metriky je jednoduché a prenechávame ho čitateľovi. ■

Definícia 4.11. Guľa $G_t(\mathbf{c})$ o strede $\mathbf{c} \in A^n$ a polomere t je množina

$$G_t(\mathbf{c}) = \{\mathbf{x} \mid \mathbf{x} \in A^n,\ d(\mathbf{x}, \mathbf{c}) \leq t\}.$$

Guľa $G_t(\mathbf{c})$ je množina všetkých takých slov, ktoré vznikli zo slova $\mathbf{c}$ nanajvýš t jednoduchými chybami.

Skúmajme, koľko prvkov obsahuje $G_t(\mathbf{c})$. Počet slov, ktoré sa líšia od $\mathbf{c} \in A^n$ práve na jednom mieste, sa rovná $n.(r-1) = \binom{n}{1}.(r-1)$, pretože na každom z n miest slova $\mathbf{c}$ zmenou pôvodného znaku na niektorý iný dostaneme $r-1$ rôznych slov, ktoré sa líšia od $\mathbf{c}$ len na jednom mieste (pripomeňme, že $|A| = r$).

Počet slov, ktoré majú od slova $\mathbf{c}$ vzdialenosť práve 2 je $\binom{n}{2}.(r-1)^2$, pretože dvojicu zmenených znakov slova $\mathbf{c}$ možno vybrať $\binom{n}{2}$ spôsobmi a každú takúto dvojicu možno nahradiť $(r-1)^2$ spôsobmi znakmi rôznymi od pôvodných.

Podobne sa ukáže, že počet slov, ktoré majú vzdialenosť od slova $\mathbf{c}$ rovnú i je $\binom{n}{i}.(r-1)^i$. Samotné slovo $\mathbf{c}$ je tiež prvkom gule $G_t(\mathbf{c})$ a prispieva k počtu jej prvkov číslom $1 = \binom{n}{0}.(r-1)^0$. Počet slov v $G_t(\mathbf{c})$ je teda

$$|G_t(\mathbf{c})| = \sum_{i=0}^{t} \binom{n}{i}.(r-1)^i\ . \tag{4.33}$$

Počet prvkov gule $G_t(\mathbf{c})$ nezávisí na tom, aké slovo $\mathbf{c}$ sme vybrali za jej stred – všetky gule o rovnakom polomere t majú rovnakú mohutnosť (4.33).

Definícia 4.12. Hovoríme, že **kód $\mathcal{K}$ opravuje t jednoduchých chýb**, ak pre slovo $\mathbf{y}$, ktoré vzniklo z niektorého kódového slova nanajvýš t jednoduchými chybami, existuje jediné slovo $\mathbf{x}$ také, že $d(\mathbf{x}, \mathbf{y}) \leq t$.

Všimnime si, že ak $\mathbf{b} \in G_t(\mathbf{c}_1) \cap G_t(\mathbf{c}_2)$, potom slovo $\mathbf{b}$ mohlo vzniknúť nanajvýš t jednoduchými chybami z oboch slov $\mathbf{c}_1$, $\mathbf{c}_2$. Ak má teda kód $\mathcal{K}$ opravovať t chýb, musí byť pre ľubovoľnú dvojicu $\mathbf{c}_1$, $\mathbf{c}_2$ rôznych kódových slov

$$G_t(\mathbf{c}_1) \cap G_t(\mathbf{c}_2) = \emptyset\ . \tag{4.34}$$

Platí aj obrátené tvrdenie. Ak pre ľubovoľnú dvojicu rôznych kódových slov kódu $\mathcal{K}$ platí (4.33), potom kód $\mathcal{K}$ opravuje t chýb.

Predpokladajme, že kód $\mathcal{K} \subseteq A^n$ opravuje t jednoduchých chýb. Keďže $|A^n| = r^n$, pre počet kódových slov $|\mathcal{K}|$ vzhľadom na (4.33) a (4.34) platí

$$\sum_{i=0}^{t} \binom{n}{i} .(r-1)^i \ . \ |\mathcal{K}| \leq r^n \ . \tag{4.35}$$

Pri návrhu kódu opravujúceho t jednoduchých chýb sa snažíme čo najlepšie využiť priestor (A^n, d). Ideálne by bolo, keby sme dostali taký systém disjunktných gúľ o polomere t, ktorý by pokrýval celú množinu A^n, t. j. keby v (4.35) platila rovnosť.

Definícia 4.13. Hovoríme, že **kód** $\mathcal{K} \subseteq A^n$ **je** t**-perfektný kód**, ak

$$\forall \mathbf{a},\ \mathbf{b} \in A^n, \quad \mathbf{a} \neq \mathbf{b} \quad G_t(\mathbf{a}) \cap G_t(\mathbf{b}) = \emptyset \ ,$$
$$\bigcup_{\mathbf{a} \in \mathcal{K}} G_t(\mathbf{a}) = A^n \ .$$

Veta 4.13. *Kód $\mathcal{K}$ opravuje t-násobné chyby práve vtedy, keď*

$$\Delta(\mathcal{K}) \geq 2t + 1 \ , \tag{4.36}$$

kde $\Delta(\mathcal{K})$ je minimálna vzdialenosť kódu $\mathcal{K}$ (pozri definíciu 4.7, vzťah (4.18) *na str. 83).*

Dôkaz. Nech platí (4.36).
Keby existovali $\mathbf{a} \in \mathcal{K}$, $\mathbf{b} \in \mathcal{K}$ také, že $G_t(\mathbf{a}) \cap G_t(\mathbf{b}) \neq \emptyset$, vezmeme $\mathbf{c} \in G_t(\mathbf{a}) \cap G_t(\mathbf{b})$. Podľa trojuholníkovej nerovnosti platí

$$d(\mathbf{a}, \mathbf{b}) \leq \underbrace{d(\mathbf{a}, \mathbf{c})}_{\leq t} + \underbrace{d(\mathbf{c}, \mathbf{b})}_{\leq t} \leq 2t,$$

čo je v spore s predpokladom, že $\Delta\mathcal{K} \geq 2t + 1$.

Nech kód $\mathcal{K} \subseteq A^n$ opravuje t jednoduchých chýb. Potom pre ľubovoľné $\mathbf{a},\ \mathbf{b} \in \mathcal{K}$ je $G_t(\mathbf{a}) \cap G_t(\mathbf{b}) = \emptyset$. Nech $d(\mathbf{a}, \mathbf{b}) = s \leq 2t$. Vytvorme postupnosť

$$\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_s \tag{4.37}$$

tak, že položíme $\mathbf{a}_0 = \mathbf{a}$, a keď už máme definované $\mathbf{a}_i$, definujeme $\mathbf{a}_{i+1}$ nasledovne: Postupne prechádzame znakmi slova $\mathbf{a}_i$ a porovnávame ich so znakmi slova $\mathbf{b}$ na rovnakej pozícii. Keď poprvýkrát natrafíme na znak slova $\mathbf{a}_i$ na pozícii k, ktorý sa nezhoduje s rovnako položeným znakom slova $\mathbf{b}$, definujeme slovo

$\mathbf{a}_{i+1}$ ako slovo $\mathbf{a}_i$ so zameneným znakom na pozícii k k-tym znakom slova $\mathbf{b}$. Postupnosť (4.37) predstavuje jeden z možných spôsobov, ako sa postupne pôsobením jednej jednoduchej chyby za druhou transformovalo slovo $\mathbf{a}$ na slovo $\mathbf{b}$.

Vidíme, že $\mathbf{a}_s = \mathbf{b}$, $d(\mathbf{a}, \mathbf{a}_i) = i$ a $d(\mathbf{a}_i, \mathbf{b}) = s - i$ pre $i = 1, 2, \ldots, s$. Preto $d(\mathbf{a}, \mathbf{a}_t) = t$, $\mathbf{a}_t \in G_t(\mathbf{a})$ a tiež $d(\mathbf{a}_t, \mathbf{b}) = s - t \leq 2t - t = t$, a teda $\mathbf{a}_t \in G_t(\mathbf{b})$, čo je v spore s predpokladom, že $G_t(\mathbf{a}) \cap G_t(\mathbf{b}) = \emptyset$. ■

Príklad 4.18. Majme abecedu $A = \{a_1, a_2, \ldots, a_r\}$. Opakovací kód dĺžky k je blokový kód, ktorého kódové slová pozostávajú z k rovnakých znakov, t. j. $\mathcal{K} = \{a_1 a_1 \ldots a_1,\ a_2 a_2 \ldots a_2,\ \ldots\ ,\ a_r a_r \ldots a_r\}$. Minimálna vzdialenosť opakovacieho kódu dĺžky k je $\Delta\mathcal{K} = k$ a takýto kód opravuje t-násobné chyby pre $t < k/2$. Špeciálne pre $r = 2$ (t. j. pre binárnu abecedu A)a k nepárne, t. j. $k = 2t + 1$ je opakovací kód t-perfektný.

Príklad 4.19. Kód s kontrolou parity (pozri príklad 4.6, str. 84) má minimálnu vzdialenosť 2, a preto neopravuje ani jednu jednoduchú chybu.

Príklad 4.20. Kód dvojrozmernej kontroly parity. Je to binárny kód, pri ktorom informačné znaky zapíšeme do matice typu (p, q). Potom ku každému riadku pridáme jeden symbol kontroly parity riadku a ku každému stĺpcu pridáme jeden symbol kontroly parity stĺpca – oba kontrolné znaky tak, aby sme dosiahli párnu paritu riadkov i stĺpcov a nakoniec pridáme znak „kontroly kontrol“ tak, aby aj parita výslednej matice bola párna. Tento kód opraví jednu jednoduchú chybu – takáto chyba zmení paritu práve jedného riadku i a práve jedného stĺpca j potom chybný znak je na mieste (i, j). Príklad kódového slova dĺžky 32 pre $p = 3$, $q = 7$:

101	0 ← kontrola parity riadku
000	0
001	1
010	1
111	1
111	1
000	0
kontroly parity stĺpcov → 110	0 ← celková kontrola parity

Majme kód $\mathcal{K}$, ktorý opravuje t chýb. Ak sme už prijali nejaké slovo $\mathbf{a}$ (či už s chybami, alebo bez nich), potrebujeme predpis, ako zo slova $\mathbf{a}$ dostať pôvodné vyslané slovo bez chýb.

Definícia 4.14. Dekódovanie kódu $\mathcal{K}$ je ľubovoľné zobrazenie δ, ktorého obor hodnôt je $\mathcal{K}$, ktorého definičný obor $\mathcal{D}(\delta)$ je podmnožinou množiny A^n všetkých slov dĺžky n a obsahuje $\mathcal{K}$ a pre ľubovoľné $\mathbf{a} \in \mathcal{K}$ je $\delta(\mathbf{a}) = \mathbf{a}$.

$$\mathcal{K} \subset \mathcal{D}(\delta) \subseteq A^n, \qquad \delta : \mathcal{D}(\delta) \to \mathcal{K}\ , \qquad \forall \mathbf{a} \in \mathcal{K}\ \ \delta(\mathbf{a}) = \mathbf{a}\ .$$

Ak $\mathcal{D}(\delta) = A^n$, hovoríme, že dekódovanie δ je **úplné**, inak hovoríme, že dekódovanie δ je **čiastočné**.

Pri definícii pojmu „dekódovanie“ prichádza k nasledujúcemu terminologickému problému: Ak je zobrazenie K kódovanie, potom dekódovaním by malo byť inverzné zobrazenie K^{-1}. Lenže pri problematike kódov opravujúcich t chýb nezávisí ani tak na tvare kódovania K, ako na vlastnostiach množiny kódových slov. Preto do úvah o objavovaní a opravovaní chýb ani konkrétny tvar kódovania nezahŕňame. Ak chceme pri prenose slov s chybami zistiť, aký znak x bol zakódovaný a poslaný, keď sme prijali slovo $\mathbf{a}$, identifikujeme znak x ako $x = K^{-1} \circ \delta(\mathbf{a})$. Určiť tvar inverzného zobrazenia k vzájomne jednoznačnému zobrazeniu K nebýva ťažké, problémom však býva určiť zobrazenie δ. Aby sme ani do ďalších úvah nemuseli zavádzať konkrétny tvar kódovania K, dohodneme sa, že pod dekódovaním budeme rozumieť zobrazenie δ podľa definície 4.14 tak, ako ho používa i väčšina literatúry o kódovaní.

U niektorých kódov môžeme rozlíšiť jednotlivé znaky na informačné a kontrolné. Kontrolné znaky sú úplne určené informačnými znakmi. Napríklad dvanásťmiestne medzinárodné číslo vagónu má prvých jedenásť znakov informačných a jeden – posledný dvanásty znak kontrolný. Podobne je to s ISBN číslom knihy, či EAN kódom tovar atď. Kód s kontrolou parity dĺžky 8 má 7 informačných znakov a jeden kontrolný znak.

Ak vieme, ako sú definované jednotlivé položky resp. znaky kódových slov, nie je problém rozdeliť znaky na informačné a kontrolné. Ako to však urobiť pre kód $\mathcal{K} \subseteq A^n$, keď poznáme len to, ako vyzerá množina kódových slov? Odpoveď dáva nasledujúca definícia:

Definícia 4.15. Nech $\mathcal{K} \subseteq A^n$ je blokový kód dĺžky n. Hovoríme, že **kód $\mathcal{K}$ má k informačných a $n-k$ kontrolných znakov**, ak existuje vzájomne jednoznačné zobrazenie $\phi : A^k \leftrightarrow \mathcal{K}$. Zobrazenie ϕ nazveme **kódovanie informačných znakov**.

Príklad 4.21. Opakovací kód dĺžky 5 s abecedou $A = \{0,1,2,3,4,5,6,7,8,9\}$ má jeden informačný znak a 4 znaky kontrolné, pretože zobrazenie ϕ definované

$$\begin{array}{lllll} \phi(0) = 00000 & \phi(1) = 11111 & \phi(2) = 22222 & \phi(3) = 33333 & \phi(4) = 44444 \\ \phi(5) = 55555 & \phi(6) = 66666 & \phi(7) = 77777 & \phi(8) = 88888 & \phi(9) = 99999 \end{array}$$

je vzájomne jednoznačné zobrazenie $\phi : A^1 \leftrightarrow \mathcal{K}$.

Príklad 4.22. Zdvojovací kód dĺžky $2n$ má n informačných a n kontrolných znakov. Kódovanie informačných znakov $\phi : A^n \leftrightarrow \mathcal{K}$ definujeme predpisom

$$\phi(a_1 a_2 \dots a_n) = a_1 a_1 a_2 a_2 \dots a_n a_n .$$

Príklad 4.23. Kód dva z piatich (pozri príklad 4.5) vôbec nemá oddelené informačné a kontrolné znaky. Počet kódových slov tohoto kódu je 10 – nie je mocninou čísla 2, a preto nemôže existovať vzájomne jednoznačné zobrazenie množiny $\{0, 1\}$ na množinu kódových slov mohutnosti 10.

V mnohých príkladoch sme videli, že kontrolná číslica bola posledným znakom kódového slova. Podobne by sme si priali, aby aj pri kódoch s k informačnými a $n-k$ kontrolnými znakmi najprv v kódovom slove vystupovali informačné a až potom kontrolné znaky. Takéto kódovanie informačných znakov nazývame systematické. Presne tento pojem určuje nasledujúca definícia:

Definícia 4.16. Blokový kód $\mathcal{K}$ je **systematický**, ak pre každé slovo $a_1 a_2 \dots a_k \in A^k$ existuje práve jedno kódové slovo $\mathbf{a} \in \mathcal{K}$ také, že

$$\mathbf{a} = a_1 a_2 \dots a_k, a_{k+1} \dots a_n \ .$$

Príklad 4.24. Opakovací kód je systematický s $k = 1$. Kód s kontrolou parity dĺžky 8 je systematický s $k = 7$. Kód medzinárodného čísla vozňa je systematický s $k = 11$.

Príklad 4.25. Zdvojovací kód s dĺžkou $2n$ väčšou ako 2 nie je systematický.

Veta 4.14. *Nech $\mathcal{K}$ je systematický kód s k informačnými a $n - k$ kontrolnými znakmi. Potom pre minimálnu vzdialenosť $\Delta\mathcal{K}$ platí*

$$\Delta\mathcal{K} \leq n - k + 1 \ . \tag{4.38}$$

Dôkaz. Zvoľme dve slová $\mathbf{a} = a_1 a_2 \dots a_{k-1} a_k \in A^k$, $\overline{\mathbf{a}} = a_1 a_2 \dots a_{k-1} \overline{a}_k \in A^k$ líšiace sa len v poslednom k-tom znaku. Pretože kód $\mathcal{K}$ je systematický, ku každému z takýchto slov existuje práve jedno slovo $\mathbf{b}$ resp. $\overline{\mathbf{b}}$ kódu $\mathcal{K}$, také, že $\mathbf{a}$ je prefixom $\mathbf{b}$, resp. $\overline{\mathbf{a}}$ je prefixom $\overline{\mathbf{b}}$:

$$\mathbf{b} = a_1 a_2 \dots a_{k-1} a_k a_{k+1} \dots a_n \ ,$$
$$\overline{\mathbf{b}} = a_1 a_2 \dots a_{k-1} \overline{a}_k \overline{a}_{k+1} \dots \overline{a}_n \ .$$

Keďže sa slová $\mathbf{b}$, $\overline{\mathbf{b}}$ zhodujú na $k-1$ miestach môžu sa nezhodovať najviac na $n-(k-1)=n-k+1$ miestach. Je $d(\mathbf{b},\overline{\mathbf{b}}) \leq n-k+1$ a teda $\Delta\mathcal{K} \leq n-k+1$. ■

Dôsledok Kód $\mathcal{K}$ s k informačnými a $n-k$ kontrolnými znakmi môže opravovať najviac $\left[\frac{n-k}{2}\right]$ chýb (kde $[x]$ je celá časť čísla x).

Príklad 4.26. Pre zdvojovací kód dĺžky $n = 2t$ je $k = t$, $n - k = t$, ale minimálna vzdialenosť tohoto kódu je 2, čo je pre veľké t hlboko pod odhadom (4.38), ktorý pre náš prípad dáva $\Delta\mathcal{K} \leq 2t - t + 1 = t + 1$.

Definícia 4.17. Nech $\mathcal{K}$ je kód s k informačnými a $n - k$ kontrolnými znakmi. Pomer

$$R = \frac{k}{n} \tag{4.39}$$

nazveme **informačný pomer**.

Pri navrhovaní samoopravných kódov sa snažíme zabezpečiť sa proti čo najväčšiemu počtu chýb, čo vedie k zvyšovaniu počtu kontrolných znakov. Druhou prirodzenou požiadavkou je dosiahnuť čo najväčší informačný pomer, čo je v rozpore so zvyšovaním počtu kontrolných znakov. Navyše na príklade 4.26 vidíme, že nie každé zvyšovanie počtu kontrolných znakov musí viesť k zväčšovaniu minimálnej vzdialenosti kódu.

4.11 Pripomenutie niektorých algebraických štruktúr

Grupa je množina G spolu s binárnou operáciou . priraďujúcou každým dvom prvkom $a \in G$, $b \in G$ prvok $a.b$ (krátko len ab) tak, že platí:

(i) $\forall a, b \in G \; ab \in G$

(ii) $\forall a, b, c \in G \; (ab)c = a(bc)$ – asociatívny zákon

(iii) $\exists 1 \in G \; \forall a \in G \quad 1a = a1 = a$ – existencia neutrálneho prvku

(iv) $\forall a \in G \; \exists a^{-1} \in G \quad aa^{-1} = a^{-1}a = 1$ – pre každý prvok grupy existuje inverzný prvok.

Grupa G je komutatívna, ak platí $\forall a, b \in G\ ab = ba$. V tomto prípade sa zvykne grupová operácia zapisovať aditívne, t. j. $a + b$ namiesto $a.b$ a neutrálny prvok sa pri aditívnom zápise označuje ako 0. Inverzný prvok k prvku a sa v komutatívnom prípade nazýva opačný prvok a označuje sa $-a$.

Teleso je množina T obsahujúca (okrem iných prvkov) prvky 0 a 1 spolu s binárnymi operáciami + a . takými, že platí:

(i) Množina T spolu s binárnou operáciou + je komutatívna grupa s neutrálnym prvkom 0.

(ii) Množina $T - \{0\}$ spolu s binárnou operáciou . je komutatívna grupa s neutrálnym prvkom 1.

(iii) $\forall a, b, c \in G \quad a(b + c) = ab + ac$ – platí distributívny zákon

Vlastnosti telesa si možno lepšie uvedomíme, ak (i), (ii), (iii) definície rozpíšeme na jednotlivé konkrétne podmienky, ktoré musí teleso spĺňať:
Teleso je množina T obsahujúca (okrem iných prvkov) prvky 0 a 1 spolu s binárnymi operáciami + a . takými, že platí:

(T1) $\forall a, b \in T\ a + b \in T$, $ab \in T$.

(T2) $\forall a, b, c \in T\ a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$ – platia asociatívne zákony.

(T3) $\forall a, b \in T\ a + b = b + a$, $ab = ba$ – platia komutatívne zákony.

(T4) $\forall a, b, c \in T\ a(b + c) = ab + ac$ – platí distributívny zákon.

(T5) $\forall a \in T\ a + 0 = a$, $a.1 = a$ – 0 je neutrálny prvok vzhľadom k operácii „+“, 1 je neutrálny prvok vzhľadom k operácii „.“.

(T6) $\forall a \in T\ \exists(-a) \in T\ a + (-a) = 0$ – ku každému prvku T existuje opačný prvok.

(T7) $\forall a \in T$, $a \neq 0\ \exists a^{-1} \in T\ a.a^{-1} = 1$ – ku každému nenulovému prvku T existuje inverzný prvok.

Komutatívny okruh s jednotkou je množina R taká, že $0 \in R$, $1 \in R$ spolu s operáciami + a ., v ktorej platia (T1) až (T6).

Príklad 4.27. Množina celých čísel spolu s operáciami + a . je komutatívnym okruhom s jednotkou.

Faktorový okruh modulo p. Majme množinu $\mathbb{Z}_p = \{0, 1, 2, \ldots, p-1\}$. Na množine $\mathbb{Z}_p$ definujeme operácie $\oplus$, $\otimes$ nasledujúcim spôsobom

$$a \oplus b = (a + b) \mod p \qquad a \otimes b = (ab) \mod p,$$

kde $n \mod p$ je zvyšok po celočíselnom delení čísla n číslom p. Ľahko sa dá ukázať, že pre ľubovoľné prirodzené číslo $p > 1$ je $\mathbb{Z}_p$ spolu s operáciami $\oplus$, $\otimes$ komutatívnym okruhom s jednotkou, t. j. spĺňa požiadavky (T1) až (T6).

Na $\mathbb{Z}_p$ sa môžeme pozerať i s nasledujúceho hľadiska. **Triedou modulo** p nazveme podmnožinu okruhu $\mathbb{Z}$ celých čísel takú, že rozdiel jej ľubovoľných dvoch prvkov je deliteľný číslom p. Triedu označujeme jej ľubovoľným reprezentantom v hranatej zátvorke. Môžeme teda triedu modulo p obsahujúcu celé číslo a definovať nasledovne:

$$[a] = \{a + pk \mid k = 0, +1, -1, +2, -2, \ldots\}$$

Ľahko sa ukáže, že ak dve triedy $[a]$, $[b]$ majú aspoň jeden spoločný prvok, potom $[a] = [b]$. Ďalej je vidieť, že všetky triedy okruhu $\mathbb{Z}$ modulo p sú $[0], [1], \ldots, [p-1]$. Ak dva prvky a, a' patria do tej istej triedy modulo p, budeme písať

$$a \equiv a' \mod p$$

a hovoriť, že prvky a, a' sú **kongruentné modulo p.**
Označme $\mathbb{Z}_p$ množinu všetkých tried modulo p, t. j. $\mathbb{Z}_p = \{[0], [1], \ldots, [p-1]\}$. Na $\mathbb{Z}_p$ definujeme operácie + a . predpisom:

$$[a] + [b] = [a + b] \qquad [a].[b] = [ab]$$

Ukazuje sa, že takto definované operácie sčítania a násobenia sú korektne definované – t. j. nezávisí na výbere reprezentanta triedy. Množina $\mathbb{Z}_p$ s takto definovanými operáciami + a . je znovu komutatívnym okruhom s jednotkou a nazývame ju **faktorovým okruhom okruhu celých čísel $\mathbb{Z}$ modulo p**, alebo len faktorovým okruhom modulo p.
Obidve reprezentácie $\mathbb{Z}_p = \{0, 1, 2, \ldots, p-1\}$, $\mathbb{Z}_p = \{[0], [1], \ldots, [p-1]\}$ sú ekvivalentné, pre ľubovoľné $a, b, c \in \{0, 1, 2, \ldots, p-1\}$ je $a \oplus b = c$ práve vtedy, keď $[a] + [b] = [c]$, $a \otimes b = c$ práve vtedy, keď $[a].[b] = [c]$; rozdiel je len v označení prvkov a operácií. Budeme preto používať jednoduchšiu prvú reprezentáciu, kde navyše budeme namiesto $\oplus$ a $\otimes$ používať + a . v prípade, že nedôjde k nedorozumeniu.

Príklad 4.28. Okruh $\mathbb{Z}_6$ bude mať nasledujúce tabuľky pre operácie sčítania a násobenia:

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	5

.	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Podľa vyššie uvedených tabuliek je $5.5 = 1$, t. j. inverzným prvkom prvku 5 je prvok 5. Prvky 2, 3, 4 vôbec nemajú inverzný prvok. Podmienka (T7) v $\mathbb{Z}_6$ nie je splnená – $\mathbb{Z}_6$ nie je telesom.

Pre potreby kódovania budú výhodné také faktorové okruhy $\mathbb{Z}$, ktoré sú telesami. Je teraz pred nami otázka, kedy je $\mathbb{Z}_p$ telesom. Odpoveď dáva nasledujúca veta.

Veta 4.15. *Faktorový okruh $\mathbb{Z}_p$ je telesom práve vtedy, keď p je prvočíslo.*

Elementárny dôkaz tejto vetu nájde čitateľ v knihe [1].

Lineárne priestory nad telesom T. Nech T je teleso. Lineárnym priestorom nad telesom T je množina $\mathcal{L}$ spolu s binárnou operáciou + (sčítanie) a skalárnou operáciou . (skalárne násobenie) takými, že platí

(L1) $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L}$ a $\forall t \in T \quad u + v \in \mathcal{L}$, $t.\mathbf{u} \in \mathcal{L}$.

(L2) $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{L} \qquad \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.

(L3) $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L} \qquad \mathbf{u} + \mathbf{b} = \mathbf{b} + \mathbf{u}$.

(L4) $\exists \mathbf{o} \in \mathcal{L}$ také, že $\forall \mathbf{u} \in \mathcal{L} \quad \mathbf{u} + \mathbf{o} = \mathbf{u}$

(L5) $\forall \mathbf{u} \in \mathcal{L} \quad \exists(-\mathbf{u}) \in \mathcal{L}$ také, že $\mathbf{u} + (-\mathbf{u}) = \mathbf{o}$

(L6) $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L}$ a $\forall t \in T \quad t.(\mathbf{u} + \mathbf{v}) = t.\mathbf{u} + t.\mathbf{v}$

(L7) $\forall \mathbf{u} \in \mathcal{L}$ a $\forall s, t \in T \quad (s.t)\mathbf{u} = s.(t.\mathbf{u})$

(L8) $\forall \mathbf{u} \in \mathcal{L}$ a $\forall s, t \in T \quad (s + t)\mathbf{u} = s.\mathbf{u} + t.\mathbf{u}$

(L9) $\forall \mathbf{u} \in \mathcal{L} \quad 1.u = u$.

Požiadavky (L1) až (L5) sú ekvivalentné s požiadavkou, aby $(\mathcal{L}, +)$ bola komutatívna grupa s neutrálnym prvkom $\mathbf{o}$. Pre lineárne priestory sa používa synonymum **vektorové priestory**, ich prvky sa volajú **vektory**.

Vektory $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ sa nazývajú **lineárne nezávislé**, ak zo vzťahu $\sum_{i=1}^{n} t_i \mathbf{u}_i = \mathbf{o}$ vyplýva $t_i = 0$ pre $i = 1, 2, \ldots, n$. Hovoríme, že lineárny priestor $\mathcal{L}$ je **konečne dimenzionálny**, ak existuje také prirodzené číslo k, že každá $k + 1$ prvková množina vektorov z $\mathcal{L}$ je lineárne závislá. V konečne dimenzionálnom priestore majú všetky maximálne lineárne nezávislé množiny vektorov rovnakú mohutnosť. Mohutnosť n maximálnej lineárne nezávislej podmnožiny $\mathcal{L}$ sa nazýva **dimenzia** lineárneho priestoru $\mathcal{Z}$ – v tomto prípade hovoríme, že priestor je n-dimenzionálny. **Báza** konečne dimenzionálneho lineárneho priestoru je ľubovoľná maximálna lineárne nezávislá množina jeho vektorov.

Lineárny priestor T^n je priestor n-prvkových postupností typu $\mathbf{u} = u_1 u_2 \ldots u_n$, kde $u_i \in T$ a kde je sčítanie a skalárne násobenie definované nasledovne:
Nech $\mathbf{u} = u_1 u_2 \ldots u_n$, $\mathbf{v} = v_1 v_2 \ldots v_n$, $t \in T$. Potom

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1)(u_2 + v_2) \ldots (u_n + v_n) \qquad t.\mathbf{u} = (tu_1)(tu_2) \ldots (tu_n).$$

Skalárny súčin vektorov $\mathbf{u} \in T^n$, $\mathbf{v} \in T^n$ je definovaný nasledovne

$$\mathbf{u} * \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Hovoríme, že vektory $\mathbf{u}$, $\mathbf{v}$ sú **ortogonálne**, ak $\mathbf{u} * \mathbf{v} = 0$.

Dôležitosť priestoru T^n vyplýva z nasledujúcej vety:

Veta 4.16. *Každý n-dimenzionálny vektorový priestor nad telesom T je izomorfný s priestorom T^n.*

V teórii lineárnych kódov sa vychádza z toho, že na kódovej abecede sú dané operácie + a . , s ktorými je táto abeceda konečným telesom. Potom sa na množinu všetkých n-znakových slov v kódovej abecede možno pozerať ako na n-dimenzionálny lineárny priestor. Videli sme už, že faktorové okruhy $\mathbb{Z}_p$ pre p prvočíslo sú telesami. Okrem toho existujú konečné telesá o p^n prvkoch (pozor, nie sú to však okruhy $\mathbb{Z}_{p^n}$). Mohutnosť kódovej abecedy je pre takéto úvahy obmedzená na čísla typu p^n, kde p je prvočíslo, t. j. 2,3,4,5,7,8,9,11,13,16,17..., ale nemôže byť 6,10,12,14,15 lebo tieto čísla nie sú mocninami prvočísel. Tieto obmedzenia však nie sú tragické, pretože najdôležitejšou kódovou abecedou je binárna abeceda, pre abecedy s väčším počtom znakov použijeme najbližšie teleso s väčším počtom prvkov s tým, že niektoré z nich na reprezentáciu znakov abecedy A nevyužijeme.

4.12 Lineárne kódy

V tejto časti budeme predpokladať, že kódová abeceda $A = \{a_1, a_2, \ldots, a_p\}$ má p prvkov, kde p je prvočíslo. Ďalej predpokladáme, že na abecede A sú definované operácie sčítania $+$ a súčinu $.$, s ktorými A vytvára teleso. Množinu A^n n-znakových slov pokladáme za lineárny priestor s obvykle definovaným súčtom a skalárnym násobkom vektorov.

Definícia 4.18. Kód $\mathcal{K}$ sa nazýva **lineárny** (n, k)**-kód**, ak je podpriestorom dimenzie k lineárneho priestoru A^n, t. j. ak $\dim(\mathcal{K}) = k$, a pre ľubovoľné $\mathbf{a}, \mathbf{b} \in \mathcal{K}$ a ľubovoľné $c \in A$ je

$$\mathbf{a} + \mathbf{b} \in \mathcal{K}, \qquad c.\mathbf{a} \in \mathcal{K}.$$

Lineárny (n, k)-kód ako k-dimenzionálny podpriestor priestoru A^n musí mať k-prvkovú bázu $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$. Potom každé kódové slovo $\mathbf{a} \in A^n$ má jednoznačné vyjadrenie

$$\mathbf{a} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_k\mathbf{b}_k, \tag{4.40}$$

kde $a_1, a_2, \ldots, a_n$ sú súradnice vektora $\mathbf{a}$ v báze $\mathbf{B}$. Ak $|A| = p$, potom na mieste každého a_i môže stáť p rôznych čísel, z čoho vyplýva, že existuje p^k rôznych k-tic $a_1, a_2, \ldots, a_k$, dosadením ktorých do (4.40) dostaneme p^k rôznych kódových slov kódu $\mathcal{K}$. Lineárny (n, k)-kód má teda p^k slov.

Všimnime si priradenie $\phi : A^k \to A^n$ definované

$$\forall(a_1a_2 \ldots a_k) \in A^k \quad \phi(a_1a_2 \ldots a_k) = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_k\mathbf{b}_k.$$

Zobrazenie ϕ je vzájomne jednoznačné zobrazenie $A^k \leftrightarrow \mathcal{K}$ a teda podľa definície 4.15 má lineárny (n, k)-kód $\mathcal{K}$ k informačných a $n - k$ kontrolných znakov. Zobrazenie ϕ je kódovanie informačných znakov.

Často bude výhodné využívať maticové zápisy, v ktorých vektory budú vystupovať ako jednoriadkové alebo jednostĺpcové matice. Dohodneme sa, že slová – t. j. vektory $\mathbf{a} \in A^n$ – budeme v maticových zápisoch vždy považovať za **stĺpcové matice**, t. j. ak $\mathbf{a} = a_1a_2 \ldots a_k$ sa vyskytne v maticovom zápise, budeme predpokladať, že

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \ldots \\ a_k \end{bmatrix}.$$

Ak budeme potrebovať vektor $\mathbf{a}$ v tvare jednoriadkovej matice, zapíšeme ho ako transponovanú maticu $\mathbf{a}^T$, t. j.

$$\mathbf{a}^T = \left[\begin{array}{cccc} a_1 & a_2 & \dots & a_k \end{array} \right] .$$

Skalárny súčin dvoch vektorov $\mathbf{u}$, $\mathbf{v} \in A^n$ môžeme považovať za súčin matíc a zapísať ako $\mathbf{u}^T.\mathbf{v}$.

Definícia 4.19. Nech $\mathcal{K}$ je lineárny (n,k)-kód, nech $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$ je ľubovoľná báza kódu $\mathcal{K}$. Nech $\mathbf{b}_i = (b_{i1}\ b_{i2} \dots b_{in})^T$ pre $i = 1, 2, \dots, k$. Potom matica

$$\mathbf{G} = \left[\begin{array}{c} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \dots \\ \mathbf{b}_k^T \end{array} \right] = \left[\begin{array}{cccc} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{array} \right] \tag{4.41}$$

typu $(k \times n)$ sa nazýva **generujúca matica kódu $\mathcal{K}$.**

Poznámka. Podľa definície 4.19 je generujúcou maticou kódu $\mathcal{K}$ každá matica, ktorej

a) každý riadok je kódovým slovom,

b) riadky sú lineárne nezávislé, takže hodnosť matice $\mathbf{G}$ sa rovná k,

c) každé kódové slovo je lineárnou kombináciou riadkov matice.

Ak teda z matice $\mathbf{G}$ vytvoríme ekvivalentnými riadkovými úpravami ekvivalentnú maticu $\mathbf{G}'$, potom aj matica $\mathbf{G}'$ je generujúcou maticou kódu $\mathcal{K}$.

Poznámka. Nech má lineárny (n,k)-kód pre bázu $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$ generujúcu maticu (4.41). Ak má slovo $\mathbf{a} = a_1 a_2 \dots a_n$ súradnice $u_1, u_2, \dots, u_k$ v báze $\mathbf{B}$, potom

$$\mathbf{a}^T = u_1\mathbf{b}_1^T + u_2\mathbf{b}_2^T + \dots + u_k\mathbf{b}_k^T = \left[\begin{array}{cccc} u_1 & u_2 & \dots & u_k \end{array} \right] . \left[\begin{array}{c} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \dots \\ \mathbf{b}_k^T \end{array} \right] ,$$

alebo po rozpísaní vektorov $\mathbf{b}_i^T$ podrobnejšie

$$\left[\begin{array}{cccc} a_1 & a_2 & \dots & a_n \end{array} \right] = \left[\begin{array}{cccc} u_1 & u_2 & \dots & u_k \end{array} \right] . \left[\begin{array}{cccc} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{array} \right] ,$$

alebo krátko

$$\mathbf{a}^T = \mathbf{u}^T.\mathbf{G} \ .$$

Príklad 4.29. Príklady lineárnych kódov.

a) Binárny kód dĺžky 4 s kontrolou parity – lineárny $(4,3)$-kód:
$\mathcal{K} \subset A^4, \quad A = \{0,1\}:$ 0000, 0011, 0101, 0110
1001, 1010, 1100, 1111
Báza: $B = \{0011, 0101, 1001\}$.

$$\text{Generujúca matica} \quad \mathbf{G} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

b) Ternárny opakovací kód dĺžky 5 – lineárny $(5,1)$-kód :
$\mathcal{K} \subset A^5, \quad A = \{0,1,2\}:$ 00000, 11111, 22222
Báza: $\{11111\}$.

$$\text{Generujúca matica} \quad \mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

c) Binárny zdvojovací kód dĺžky 6 – lineárny $(6,3)$-kód :
$\mathcal{K} \subset A^6, \quad A = \{0,1\}:$ 000000, 000011, 001100, 001111
110000, 110011, 111100, 111111
Báza: $\{000011, 001100, 110000\}$.

$$\text{Generujúca matica} \quad \mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

d) Dekadický kód dĺžky n s kontrolnou číslicou modulo 10 nie je lineárnym kódom, lebo neexistuje konečné teleso s počtom prvkov 10.

Definícia 4.20. Hovoríme, že dva blokové kódy $\mathcal{K}, \mathcal{K}'$ dĺžky n sú **ekvivalentné**, ak existuje permutácia π množiny $\{1, 2, \ldots, n\}$ taká, že platí

$$\forall a_1 a_2 \ldots a_n \in A^n \quad a_1 a_2 \ldots a_n \in \mathcal{K} \quad \text{práve vtedy, keď} \quad a_{\pi[1]} a_{\pi[2]} \ldots a_{\pi[n]} \in \mathcal{K}' \ .$$

Podľa definície 4.16 (str. 104) je blokový kód $\mathcal{K}$ s k informačnými a $n-k$ kontrolnými znakmi systematický, ak ku každému $a_1 a_2 \ldots a_k \in A^k$ existuje práve jedno kódové slovo $\mathbf{a} \in \mathcal{K}$ s prefixom $a_1 a_2 \ldots a_k \in A^k$. Ako sme už

ukázali, lineárny (n,k)-kód je kódom s k informačnými a $n-k$ kontrolnými znakmi, avšak nemusí byť systematický. Zdvojovací kód je $n=2k$ je lineárny kód, ktorý nie je systematický, ak $k>1$. Stačí však zmeniť poradie znakov v slove $a_1a_2,\dots a_n$ – dať najprv znaky na nepárnych miestach a potom znaky na párnych miestach a takto získaný nový kód je už systematický. Toto sa dá urobiť s každým lineárnym (n,k)-kódom.

Veta 4.17. *Lineárny (n,k)-kód $\mathcal{K}$ je systematický práve vtedy, keď k nemu existuje generujúca matica $\mathbf{G}$ typu:*

$$\mathbf{G}=\left[\begin{array}{c|c}\mathbf{E} & \mathbf{B}\end{array}\right]=\begin{bmatrix}1&0&0&\dots&0&b_{11}&b_{12}&\dots&h_{1n-k}\\0&1&0&\dots&0&b_{21}&b_{22}&\dots&b_{2n-k}\\ \dots&\dots&\dots&\dots&\dots&\dots&\dots&\dots&\dots\\0&0&0&\dots&1&b_{k1}&h_{k2}&\dots&h_{kn-k}\end{bmatrix}. \tag{4.42}$$

Dôkaz. Nech (4.42) je generujúcou maticou pre kód $\mathcal{K}$. Nech $\mathbf{u}=u_1,u_2,\dots u_k$ sú súradnice slova $\mathbf{a}=a_1a_2\dots a_n\in\mathcal{K}$ v báze, ktorú tvoria riadky generujúcej matice $\mathbf{G}$. Potom podľa poznámky 4.19 je $\mathbf{a}^T=\mathbf{b}^T.\mathbf{G}$. Špeciálne pre $\mathbf{u}=a_1a_2\dots a_k$ je

$$\mathbf{u}^T.\mathbf{G}=\begin{bmatrix}a_1&a_2&\dots&a_k\end{bmatrix}.\begin{bmatrix}1&0&0&\dots&0&b_{11}&b_{12}&\dots&b_{1n-k}\\0&1&0&\dots&0&b_{21}&b_{22}&\dots&b_{2n-k}\\ \dots&\dots&\dots&\dots&\dots&\dots&\dots&\dots&\dots\\0&0&0&\dots&1&b_{k1}&b_{k2}&\dots&b_{kn-k}\end{bmatrix}=$$
$$=\begin{bmatrix}a_1&a_2&\dots&a_k&v_{k+1}&\dots&v_n\end{bmatrix},$$

kde v_{k+i} je jednoznačne určené vzťahom:

$$v_{k+i}=\begin{bmatrix}a_1&a_2&\dots&a_k\end{bmatrix}.\begin{bmatrix}b_{1i}\\b_{2i}\\\dots\\b_{ki}\end{bmatrix}.$$

Pre každé $a_1a_2\dots a_k\in A^k$ existuje práve jedno slovo kódu $\mathcal{K}$ s prefixom $a_1a_2\dots a_k$. Kód $\mathcal{K}$ je teda systematický.

Nech je kód $\mathcal{K}$ systematický. Ak sú prvé k stĺpce generujúcej matice $\mathbf{G}$ lineárne nezávislé, riadkovými ekvivalentnými úpravami ju môžeme previesť na ekvivalentnú maticu $\mathbf{G}'$ v tvare $\mathbf{G}'=\left[\begin{array}{c|c}\mathbf{E} & \mathbf{B}\end{array}\right]$, ktorá je tiež generujúcou maticou kódu $\mathcal{K}$.

Nech teda prvé k stĺpce generujúcej matice $\mathbf{G}$ systematického kódu $\mathcal{K}$ nie sú lineárne závislé. Potom ju môžeme ekvivalentnými riadkovými úpravami transformovať na ekvivalentný tvar

$$\mathbf{G}' = \left[\begin{array}{ccccc|cccc} d_{11} & d_{12} & d_{13} & \dots & d_{1k} & d_{1(k+1)} & d_{1(k+2)} & \dots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2k} & d_{2(k+1)} & d_{2(k+2)} & \dots & d_{2n} \\ \multicolumn{5}{c|}{\dots\dots\dots\dots\dots} & \multicolumn{4}{c}{\dots\dots\dots\dots\dots} \\ d_{(k-1)1} & d_{(k-1)2} & d_{(k-1)3} & \dots & d_{(k-1)k} & d_{(k-1)(k+1)} & d_{(k-1)(k+2)} & \dots & d_{(k-1)n} \\ 0 & 0 & 0 & \dots & 0 & d_{k(k+1)} & d_{k(k+2)} & \dots & d_{kn} \end{array}\right]$$

Matica $\mathbf{G}'$ má hodnosť k, pretože je ekvivalentná s maticou $\mathbf{G}$, ktorá mala k lineárne nezávislých riadkov. Pre $\mathbf{u}$, $\mathbf{v} \in A^k$ také, že $\mathbf{u} \neq \mathbf{v}$ je $\mathbf{u}^T.\mathbf{G}' \neq \mathbf{v}^T.\mathbf{G}'$. Všimnime si, že prvých k súradníc vektora $\mathbf{u}^T.\mathbf{G}$ nezávisí na k-tej súradnici vektora $\mathbf{u}$, z čoho vyplýva, že existuje niekoľko kódových slov kódu $\mathcal{K}$ s rovnakým prefixom a teda kód $\mathcal{K}$ nie je systematický. Z predpokladu, že prvé k stĺpce generujúcej matice sú závislé, sme dostali spor. ■

Dôsledok. Lineárny (n, k)-kód $\mathcal{K}$ je systematický práve vtedy, keď jeho ľubovoľná generujúca matica $\mathbf{G}$ má prvé k stĺpce lineárne nezávislé.

Veta 4.18. *Každý lineárny (n, k)-kód $\mathcal{K}$ je ekvivalentný so systematickým lineárnym kódom.*

Dôkaz. Nech $\mathbf{G}$ je generujúca matica (n, k) kódu $\mathcal{K}$. Matica $\mathbf{G}$ má k lineárne nezávislých riadkov, a preto musí mať (aspoň jednu) k-ticu lineárne nezávislých stĺpcov. Ak prvých k stĺpcov matice $\mathbf{G}$ je lineárne nezávislých, podľa dôsledku vety 4.17 je kód $\mathcal{K}$ systematický.

Ak sú prvé k stĺpce lineárne závislé, urobíme takú permutáciu π stĺpcov, aby prvé k stĺpce boli lineárne nezávislé. Potom už príslušný kód $\mathcal{K}'$, ktorý dostaneme rovnakou permutáciou π znakov kódu $\mathcal{K}$, bude systematický. ■

Existuje i iný spôsob charakterizácie lineárneho (n, k)-kódu a to tak, že vlastnosti kódových slov charakterizujem sústavou lineárnych rovníc, ktoré musia všetky kódové slová spĺňať. Tak napríklad binárny kód dĺžky n s kontrolou parity charakterizujeme rovnicou:

$$x_1 + x_2 + \cdots + x_n = 0$$

Zdvojovací kód dĺžky $n = 2k$ charakterizujeme sústavou rovníc:

$$x_1 - x_2 = 0$$
$$x_3 - x_4 = 0$$

$$\dots$$
$$x_{2i-1} - x_{2i} = 0$$
$$\dots$$
$$x_{n-1} - x_n = 0$$

Sústava rovníc pre opakovací kód dĺžky n bude:

$$x_1 - x_2 = 0$$
$$x_1 - x_3 = 0$$
$$\dots$$
$$x_1 - x_n = 0$$

Definícia 4.21. Kontrolná matica lineárneho kódu $\mathcal{K}$ je taká matica $\mathbf{H}$ prvkov kódovej abecedy A, pre ktorú platí: Slovo $\mathbf{v} = v_1 v_2 \dots v_n$ je kódové práve vtedy, keď

$$\mathbf{H}.\mathbf{v} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{bmatrix} . \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} = \mathbf{o} \ . \tag{4.43}$$

Stručnejšie: $\mathbf{v} \in \mathcal{K}$ práve vtedy, keď $\mathbf{H}.\mathbf{v} = \mathbf{o}$.

Majme lineárny (n, k)-kód $\mathcal{K}$ s generujúcou maticou

$$\mathbf{G} = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \dots \\ \mathbf{b}_k^T \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{bmatrix} \tag{4.44}$$

typu $(k \times n)$. Aká má byť kontrolná matica kódu $\mathcal{K}$ t. j. matica $\mathbf{H}$ taká, že $\mathbf{H}.\mathbf{u} = \mathbf{o}$ práve vtedy, keď $\mathbf{u} \in \mathcal{K}$? Prvé, čo o matici $\mathbf{H}$ vieme povedať, je, že má mať n stĺpcov (už len preto, aby $\mathbf{H}.\mathbf{u}$ bolo definované pre $\mathbf{u} \in A^n$). Množina všetkých $\mathbf{u} \in A^n$ takých, že $\mathbf{H}.\mathbf{u} = \mathbf{o}$ je podpriestor priestoru A^n dimenzie rovnej $n - \dim(\mathbf{H}) = \dim(\mathcal{K}) = k$, odkiaľ $\dim(\mathbf{H}) = n - k$. Stačí teda hľadať maticu $\mathbf{H}$ ako maticu typu $((n-k) \times n)$ s $n-k$ lineárne nezávislými riadkami. Nech $\mathbf{h}^T$ je riadok matice $\mathbf{H}$. Potom pre každé kódové slovo $\mathbf{u} \in \mathcal{K}$ musí byť

$$\mathbf{u}^T.\mathbf{h} = u_1 h_1 + u_2 h_2 + \dots + u_n h_n = 0 \ . \tag{4.45}$$

Mohli by sme teda zostaviť sústavu $p^k = |\mathcal{K}|$ lineárnych rovníc typu (4.45), kde by $\mathbf{u}$ prebiehalo všetky kódové slová kódu $\mathcal{K}$. Takýto systém lineárnych rovníc by však obsahoval príliš veľa lineárne závislých rovníc. Stačí totiž, aby (4.45) platilo pre všetky prvky bázy podpriestoru $\mathcal{K}$, potom bude (4.45) platiť aj pre všetky prvky podpriestoru $\mathcal{K}$. Pre $\mathbf{h}$ možno zostaviť túto sústavu lineárnych rovníc:

$$\left.\begin{array}{rcl} \mathbf{b}_1^T.\mathbf{h} & = & 0 \\ \mathbf{b}_2^T.\mathbf{h} & = & 0 \\ \cdots & & \\ \mathbf{b}^T.\mathbf{h} & = & 0 \end{array}\right\},$$

čo sa dá v maticovom tvare zapísať

$$\mathbf{G}.\mathbf{h} = \mathbf{o}\ . \tag{4.46}$$

Keďže dimenzia matice $\mathbf{G}$ je k, množina všetkých riešení sústavy (4.46) je podpriestor dimenzie $(n-k)$ a preto možno nájsť $(n-k)$ lineárne nezávislých riešení sústavy (4.46) $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{n-k}$, ktoré budú riadkami hľadanej kontrolnej matice $\mathbf{H}$, t. j.

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \cdots \\ \mathbf{h}_{n-k}^T \end{bmatrix}.$$

Všimnime si, že

$$\mathbf{G}.\mathbf{H}^T = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \cdots \\ \mathbf{b}_k^T \end{bmatrix}_{k\times n} \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \ldots & \mathbf{h}_{n-k} \end{bmatrix}_{n\times(n-k)} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{bmatrix}_{k\times(n-k)}.$$

Majme maticu $\mathbf{H}$ typu $((n-k)\times n)$ dimenzie $\dim(\mathbf{H}) = (n-k)$ pre ktorú platí $\mathbf{G}.\mathbf{H}^T = \mathbf{O}_{k\times(n-k)}$, kde $\mathbf{O}_{k\times(n-k)}$ je nulová matica typu $(k\times(n-k))$. Označme $\mathcal{N} \subseteq A^n$ priestor všetkých riešení rovnice $\mathbf{H}\mathbf{u} = \mathbf{o}$. Keďže pre všetky prvky bázy kódu $\mathbf{K}$ platí $\mathbf{H}.\mathbf{b}_i = \mathbf{o}$, $i = 1, 2, \ldots k$, platí aj pre ľubovoľné kódové slovo $\mathbf{u} \in \mathcal{K}$, $\mathbf{u} = \sum_{i=1}^{k} u_i\mathbf{b}_i$:

$$\mathbf{H}.\mathbf{u} = \mathbf{H}.\sum_{i=1}^{k} u_i\mathbf{b}_i = \sum_{i=1}^{k} \mathbf{H}.(u_i\mathbf{b}_i) = \sum_{i=1}^{k} u_i(\mathbf{H}.\mathbf{b}_i) = \sum_{i=1}^{k} u_i.\mathbf{o} = \mathbf{o}\ .$$

Máme teda $\mathcal{K} \subseteq \mathcal{N}$. Pretože $\dim(\mathbf{H}) = (n-k)$, $\dim(\mathcal{N})$ sa rovná $n - \dim(\mathbf{H}) = k$. Keďže $\mathcal{K} \subseteq \mathcal{N}$ je báza $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$ bázou priestoru $\mathcal{N}$, a teda $\mathcal{K} = \mathcal{N}$.

Práve dokázané skutočnosti môžeme sformulovať do nasledujúcej vety.

Veta 4.19. *Nech $\mathcal{K}$ je lineárny (n,k)-kód s generujúcou maticou $\mathbf{G}$ typu $(k \times n)$. Potom matica $\mathbf{H}$ typu $((n-k) \times n)$ je kontrolnou maticou kódu $\mathcal{K}$ práve vtedy, keď*

$$\dim(\mathbf{H}) = (n-k) \quad a \quad \mathbf{G}.\mathbf{H}^T = \mathbf{O}_{k\times(n-k)}, \tag{4.47}$$

kde $\mathbf{O}_{k\times(n-k)}$ je nulová matica typu $(k \times (n-k))$.

Pre systematické kódy je situácie jednoduchšia, ako hovorí nasledujúca veta.

Veta 4.20. *Lineárny (n,k)-kód $\mathcal{K}$ s generujúcou maticou $\mathbf{G} = \left[\begin{array}{c|c} \mathbf{E}_{k\times k} & \mathbf{B} \end{array}\right]$ má kontrolnú maticu $\mathbf{H} = \left[\begin{array}{c|c} -\mathbf{B}^T & \mathbf{E}_{(n-k)\times(n-k)} \end{array}\right]$.*

Dôkaz. Označme $m = n-k$. Potom môžeme matice $\mathbf{G}$, $\mathbf{H}$ rozpísať nasledovne:

$$\mathbf{G} = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \ldots \\ \mathbf{b}_p^T \\ \ldots \\ \mathbf{b}_k^T \end{bmatrix} = \left[\begin{array}{cccccc|cccccc} 1 & 0 & \ldots & 0 & \ldots & 0 & b_{11} & b_{12} & \ldots & b_{1q} & \ldots & b_{1m} \\ 0 & 1 & \ldots & 0 & \ldots & 0 & b_{21} & b_{22} & \ldots & b_{1q} & \ldots & b_{2m} \\ \multicolumn{6}{c|}{\ldots} & & & & & & \\ 0 & 0 & \ldots & 1 & \ldots & 0 & b_{p1} & b_{p2} & \ldots & b_{pq} & \ldots & b_{pm} \\ \multicolumn{6}{c|}{\ldots} & & & & & & \\ 0 & 0 & \ldots & 0 & \ldots & 1 & b_{k1} & b_{k2} & \ldots & b_{kq} & \ldots & b_{km} \end{array}\right],$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \ldots \\ \mathbf{h}_q^T \\ \ldots \\ \mathbf{h}_m^T \end{bmatrix} = \left[\begin{array}{cccccc|cccccc} -b_{11} & -b_{21} & \ldots & -b_{p1} & \ldots & -b_{k1} & 1 & 0 & \ldots & 0 & \ldots & 0 \\ -b_{12} & -b_{22} & \ldots & -b_{p2} & \ldots & -b_{k2} & 0 & 1 & \ldots & 0 & \ldots & 0 \\ \multicolumn{6}{c|}{\ldots} & & & & & & \\ -b_{1q} & -b_{2q} & \ldots & -b_{pq} & \ldots & -b_{kq} & 0 & 0 & \ldots & 1 & \ldots & 0 \\ \multicolumn{6}{c|}{\ldots} & & & & & & \\ -b_{1m} & -b_{2m} & \ldots & -b_{pm} & \ldots & -b_{km} & 0 & 0 & \ldots & 0 & \ldots & 1 \end{array}\right].$$

Pre $\mathbf{b}_p$, $\mathbf{h}_q$ platí

$$\begin{array}{rcl} \mathbf{b}_p^T & = & [\ 0 \quad 0 \quad \ldots \quad 1 \quad \ldots \quad 0 \quad b_{p1} \quad b_{p2} \quad \ldots \quad b_{pq} \quad \ldots \quad b_{pm}\ ] \\ \mathbf{h}_q^T & = & [\ -b_{1q} \quad -b_{2q} \quad \ldots \quad -b_{pq} \quad \ldots \quad -b_{kq} \quad 0 \quad 0 \quad \ldots \quad 1 \quad \ldots \quad 0\ ] \end{array}$$

a preto je $\mathbf{b}_p^T.\mathbf{h}_q = (-b_{pq} + b_{pq}) = 0$ pre každé $p,\ q \in \{1, 2, \ldots, n\}$, z čoho

$$\mathbf{G}.\mathbf{H}^T = \mathbf{O}_{k\times(n-k)}.$$

Keďže matica $\mathbf{H}$ s $m = n-k$ riadkami obsahuje jednotkovú podmaticu $\mathbf{E}_{(n-k)\times(n-k)}$, je $\dim(H) = n-k$. Podľa vety je 4.19 je $\mathbf{H}$ kontrolnou maticou kódu $\mathcal{K}$.

Definícia 4.22. Nech $\mathcal{K} \subseteq A^n$ je lineárny (n,k) kód. **Duálny kód** $\mathcal{K}^\perp$ kódu $\mathcal{K}$ definujeme ako

$$\mathcal{K}^\perp = \{\mathbf{v} \mid \mathbf{a}.\mathbf{v} = 0 \ \forall \mathbf{a} \in \mathcal{K}\}.$$

Veta 4.21. *Nech $\mathcal{K} \subseteq A^n$ je lineárny (n,k)-kód s generujúcou maticou $\mathbf{G}$ a kontrolnou maticou $\mathbf{H}$. Potom duálny kód $\mathcal{K}^\perp$ kódu $\mathcal{K}$ je lineárny $(n, n-k)$-kód s generujúcou maticou $\mathbf{H}$ a kontrolnou maticou $\mathbf{G}$.*

Dôkaz. Platí $\mathbf{v} \in \mathcal{K}^\perp$ práve vtedy, keď

$$\mathbf{G}.\mathbf{v} = \mathbf{o}. \tag{4.48}$$

Pretože $\mathcal{K}^\perp$ je množinou riešení rovnice (4.48) a $\dim(\mathbf{G}) = k$, je $\mathcal{K}^\perp$ $(n-k)$-dimenzionálnym podpriestorom A^n – t. j. lineárnym $(n,(n-k))$-kódom s kontrolnou maticou $\mathbf{G}$.

Pretože $\mathbf{H}.\mathbf{G}^T = \left((\mathbf{G}^T)^T.\mathbf{H}^T\right)^T = \left(\mathbf{G}.\mathbf{H}^T\right)^T = \mathbf{O}^T_{k\times(n-k)} = \mathbf{O}_{(n-k)\times k}$, je každý riadok matice $\mathbf{H}$ ortogonálny s podpriestorom $\mathcal{K}$ a teda kódovým slovo kódu $\mathcal{K}^\perp$. Pretože $\dim(\mathbf{H}) = (n-k)$, generujú riadky matice $\mathbf{H}$ celý priestor $\mathcal{K}^\perp$, t. j. matica $\mathbf{H}$ je generujúcou maticou kódu $\mathcal{K}^\perp$.

Príklad 4.30. Duálny kód binárneho opakovacieho kódu $\mathcal{K}$ dĺžky 5 je kód obsahujúci všetky binárne slová $v_1v_2\ldots v_n$ také, že

$$v_1 + v_2 + v_3 + v_4 + v_5 = 0.$$

Kód $\mathcal{K}^\perp$ je kód kontroly paritou.

Príklad 4.31. Duálny kód k binárnemu zdvojovaciemu kódu $\mathcal{K}$ je samotný kód $\mathcal{K}$, teda $\mathcal{K}^\perp = \mathcal{K}$. Ten má totiž generujúcu maticu

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Ľahko sa presvedčíme, že $\mathbf{G}.\mathbf{G}^T = \mathbf{O}_{3\times 3}$ – generujúca matica binárneho zdvojovacieho kódu $\mathcal{K}$ je zároveň i jeho kontrolnou maticou.

4.13 Lineárne kódy a objavovanie chýb

V časti 4.7 v definícii 4.7 sme všeobecne definovali, čo to znamená, že kód objavuje niekoľkonásobné chyby – totiž kód $\mathcal{K}$ objavuje t-násobné jednoduché

chyby, ak pri zmene ľubovoľných t znakov ľubovoľného kódového slova $\mathbf{u}$ vznikne nekódové slovo.

Teória lineárnych kódov nám umožňuje podrobnejšie modelovať mechanizmus vzniku niekoľkonásobnej chyby a to tak, ako keby sa k vyslanému slovu $\mathbf{v} = v_1 v_2 \dots v_n$ behom prenosu pripočítalo slovo $\mathbf{e} = e_1 e_2 \dots e_n$. Potom namiesto slova $\mathbf{v}$ prijmeme slovo $\mathbf{w} = w_1 w_2 \dots w_n$ pre ktoré platí $\mathbf{w} = \mathbf{v} + \mathbf{e}$. Slovo $\mathbf{e}$ nazývame **chybové slovo**.

Definícia 4.23. Hovoríme, že lineárny kód $\mathcal{K}$ **objavuje chybové slovo e**, ak pre každé kódové slovo $\mathbf{v}$ je slovo $\mathbf{v} + \mathbf{e}$ nekódovým slovom.

Definícia 4.24. Hammingova váha $\|\mathbf{a}\|$ slova $\mathbf{a} \in A^n$ je počet nenulových znakov slova $\mathbf{a}$.

Veta 4.22. *Každý binárny lineárny kód obsahuje buď len slová párnej váhy, alebo má rovnaký počet slov párnej a nepárnej váhy.*

Dôkaz. Skutočne, ak existuje kódové slovo $\mathbf{v}$ nepárnej váhy, fixujme ho a definujme zobrazenie $f : \mathcal{K} \to \mathcal{K}$ predpisom

$$f(\mathbf{w}) = \mathbf{w} + \mathbf{v} \ .$$

Je ihneď vidieť, že f je vzájomne jednoznačné zobrazenie $\mathcal{K}$ na $\mathcal{K}$ priraďujúce každému slovu párnej váhy slovo nepárnej váhy a naopak. Z toho už vyplýva, že počet slov párnej váhy sa rovná počtu slov nepárnej váhy. ■

Všimnime si, že lineárny kód objavuje t-násobné chyby práve vtedy, keď objavuje všetky chybové slová Hammingovej váhy menšej alebo rovnej t.

Pre objavovanie a opravovanie chýb má podstatný význam minimálna vzdialenosť $\Delta(\mathcal{K})$ blokového kódu $\mathcal{K}$, ktorá bola v definícii 4.7 na str. 83 definovaná ako minimum z Hammingových vzdialeností všetkých dvojíc nerovnakých slov kódu $\mathcal{K}$. Ak totiž $d = \Delta(\mathcal{K})$, kód $\mathcal{K}$ objavuje všetky $(d-1)$-násobné chyby a opravuje všetky t-násobné chyby pre $t < \dfrac{d}{2}$ (pozri vetu 4.13 na str. 101).

Pre lineárny kód sa $\Delta(\mathcal{K})$ určí ešte jednoduchšie.

Veta 4.23. *Pre lineárny kód $\mathcal{K}$ sa minimálna vzdialenosť kódu $\Delta(\mathcal{K})$ rovná minimu z Hammingových váh všetkých nenulových slov kódu $\mathcal{K}$, t. j.*

$$\Delta(\mathcal{K}) = \min_{\mathbf{u} \in \mathcal{K}, \mathbf{u} \neq \mathbf{o}} \{\|\mathbf{u}\|\} \ .$$

Dôkaz. 1. Majme $\mathbf{u},\ \mathbf{v} \in \mathcal{K}$ také, že $d(\mathbf{u},\mathbf{v}) = \Delta(\mathcal{K})$. Nech $\mathbf{w} = \mathbf{u} - \mathbf{v}$. Slovo $\mathbf{w}$ má práve toľko nenulových znakov, v koľkých znakoch sa líšia slová $\mathbf{u},\ \mathbf{v}$, preto je

$$\min_{\mathbf{u}\in\mathcal{K},\mathbf{u}\neq\mathbf{o}}\{\|\mathbf{u}\|\} \leq \|\mathbf{w}\| = d(\mathbf{u},\mathbf{v}) = \Delta(\mathcal{K}) \ . \tag{4.49}$$

2. Vezmime $\mathbf{w} \in \mathcal{K}$ také, že $\|w\| = \min_{\mathbf{u}\in\mathcal{K},\mathbf{u}\neq\mathbf{o}}\{\|\mathbf{u}\|\}$. Potom

$$\Delta(\mathcal{K}) \leq d(\mathbf{o},\mathbf{v}) = \|\mathbf{w}\| \leq \min_{\mathbf{u}\in\mathcal{K},\mathbf{u}\neq\mathbf{o}}\{\|\mathbf{u}\|\} \ . \tag{4.50}$$

Vzťahy (4.49) a (4.50) už dávajú tvrdenie vety. ■

Definícia 4.25. Nech $\mathbf{H}$ je kontrolná matica lineárneho kódu $\mathcal{K}$, nech $\mathbf{v} = v_1 v_2 \dots v_n \in A^n$ je ľubovoľné slovo abecedy A dĺžky n. **Syndróm slova v** je slovo $\mathbf{s} = s_1 s_2 \dots s_n$, pre ktoré platí

$$\mathbf{H}.\begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ s_n \end{bmatrix}, \qquad \text{skrátene} \quad \mathbf{H}.\mathbf{v} = \mathbf{s} \ .$$

Ak teda prijmeme slovo $\mathbf{w}$, vypočítame jeho syndróm $\mathbf{s} = \mathbf{H}\mathbf{w}$, a ak $\mathbf{s} \neq \mathbf{o}$, vieme, že došlo k chybe. Navyše vieme, že syndróm prijatého slova $\mathbf{w} = \mathbf{v} + \mathbf{e}$ je rovnaký, ako syndróm chybového slova $\mathbf{e}$. Je totiž

$$\mathbf{H}\mathbf{w} = \mathbf{H}(\mathbf{v} + \mathbf{e}) = \mathbf{H}\mathbf{v} + \mathbf{H}\mathbf{e} = \mathbf{o} + \mathbf{H}\mathbf{e} = \mathbf{H}\mathbf{e} \ .$$

Keďže kód $\mathcal{K}$ je podpriestor práve všetkých riešení rovnice $\mathbf{H} = \mathbf{o}$, rovnica $\mathbf{H}\mathbf{e} = \mathbf{s}$ má množinu všetkých riešení v tvare $\mathbf{e} + \mathbf{k}$, kde $\mathbf{k} \in \mathcal{K}$. Túto množinu budeme v ďalšom označovať ako $\mathbf{e} + \mathcal{K}$.

Veta 4.24. *Nech $\mathcal{K}$ je lineárny kód s kontrolnou maticou $\mathbf{H}$. Nech d je minimum z počtu lineárne závislých stĺpcov[2] kontrolnej matice $\mathbf{H}$. Potom pre minimálnu vzdialenosť $\Delta(\mathcal{K})$ kódu $\mathcal{K}$ platí*

$$d = \Delta(\mathcal{K}) \ .$$

[2] V kontrolnej matici $\mathbf{H}$ existuje d lineárne závislých stĺpcov, ale každých $d-1$ stĺpcov kontrolnej matice je už lineárne nezávislých.

Dôkaz. Podľa vety 4.23 sa $\Delta(\mathcal{K})$ rovná minimálnej váhe nenulového kódového slova. Nech d je minimum počtu lineárne závislých stĺpcov kontrolnej matice $\mathbf{H}$. Označme $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ stĺpce kontrolnej matice $\mathbf{H}$, t. j.

$$\mathbf{H} = \left[\begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{array} \right] .$$

Nech $\mathbf{u} \in \mathcal{K}$ je nenulové slovo s najmenšou Hammingovou váhou $\|\mathbf{u}\| = t$. Slovo $\mathbf{u}$ má na miestach $i_1, i_2, \ldots, i_t$ znaky $u_{i_1}, u_{i_2}, \ldots, u_{i_t}$ a na ostatných miestach znak 0, t. j.

$$\mathbf{u}^T = \left[\begin{array}{ccccccccccccccc} 0 & 0 & \ldots & 0 & u_{i_1} & 0 & \ldots & 0 & u_{i_2} & 0 & \ldots & \ldots & 0 & u_{i_t} & 0 & \ldots & 0 & 0 \end{array} \right] .$$

Pretože $\mathbf{u}$ je kódové slovo, je $\mathbf{Hv} = \mathbf{o}$, t. j.

$$\mathbf{Hu} = \sum_{i=1}^{n} u_i.\mathbf{c}_i = u_{i_1}\mathbf{c}_{i_1} + u_{i_2}\mathbf{c}_{i_2} + \cdots + u_{i_t}\mathbf{u}_{i_t} = \mathbf{o} \ . \tag{4.51}$$

Pretože všetky koeficienty u_{i_j} sú nenulové, sú stĺpce $\mathbf{c}_{i_1}, \mathbf{c}_{i_2}, \ldots, \mathbf{c}_{i_t}$ lineárne závislé, a preto

$$d \leq \Delta(\mathcal{K}) \ . \tag{4.52}$$

Majme d lineárne závislých stĺpcov $\mathbf{c}_{i_1}, \mathbf{c}_{i_2}, \ldots, \mathbf{c}_{i_d}$. Potom existujú čísla $u_{i_1}, u_{i_2}, \ldots, u_{i_d}$ také, že aspoň jedno z nich je nenulové a

$$u_{i_1}\mathbf{c}_{i_1} + u_{i_2}\mathbf{c}_{i_2} + \cdots + u_{i_d}\mathbf{c}_{i_d} = \mathbf{o} \ .$$

Definujme slovo $\mathbf{u}$ také, že na miestach $i_1, i_2, \ldots, i_d$ bude mať znaky $u_{i_1}, u_{i_2}, \ldots, u_{i_d}$ a na ostatných miestach znak 0, t. j.

$$\mathbf{u}^T = \left[\begin{array}{ccccccccccccccc} 0 & 0 & \ldots & 0 & u_{i_1} & 0 & \ldots & 0 & u_{i_2} & 0 & \ldots & \ldots & 0 & u_{i_t} & 0 & \ldots & 0 & 0 \end{array} \right] .$$

Potom

$$\mathbf{Hu} = \sum_{i=1}^{n} u_i.\mathbf{c}_i = u_{i_1}\mathbf{c}_{i_1} + u_{i_2}\mathbf{c}_{i_2} + \cdots + u_{i_t}\mathbf{c}_{i_d} = \mathbf{o} \ , \tag{4.53}$$

a teda $\mathbf{u}$ je nenulovým kódovým slovom, pre ktorého Hammingovu váhu platí $\|\mathbf{u}\| \leq d$ a teda

$$\Delta(\mathcal{K}) \leq d \ .$$

Posledná nerovnosť spolu s (4.52) už dáva požadované tvrdenie vety. ■

Veta 4.25. *Lineárny kód objavuje t-násobné chyby práve vtedy, keď každých t stĺpcov kontrolnej matice je lineárne nezávislých.*

Dôkaz. Označme $d = \Delta(\mathcal{K})$. Podľa predchádzajúcej vety 4.24 existuje v kontrolnej matici $\mathbf{H}$ kódu $\mathcal{K}$ d lineárne závislých stĺpcov, a pre každé $t < d$ je ľubovoľných t stĺpcov matice $\mathbf{H}$ lineárne nezávislých.

Ak kód $\mathcal{K}$ objavuje t-násobné chyby, potom musí byť $t < d$, a podľa vety 4.24 je každých t stĺpcov matice $\mathbf{H}$ lineárne nezávislých.

Ak je každých t stĺpcov matice $\mathbf{H}$ lineárne nezávislých, potom podľa vety 4.24 je $t < d$, a preto kód $\mathcal{K}$ objavuje t chýb. ■

4.14 Štandardné dekódovanie

V predchádzajúcej časti sme ukázali, ako určíme najväčší počet jednoduchých chýb, ktoré je kód $\mathcal{K}$ schopný objaviť, a akým spôsobom zistíme, či nastalo niekoľko jednoduchých chýb (pochopiteľne za predpokladu, že ich je najviac t). Ak už prijmeme nekódové slovo, chceme mu priradiť kódové slovo, ktorého pokazením toto slovo pravdepodobne vzniklo (zase za predpokladov, že počet chýb neprekročil hodnotu t). Na to slúži dekódovanie δ definované v časti 4.10 v definícii 4.14 (str. 103) ako funkcia, ktorá ma za definičný obor A^n alebo jeho časť obsahujúcu kód $\mathcal{K}$, a ktorá každému slovu zo svojho definičného oboru priraďuje kódové slovo, pričom je δ na $\mathcal{K}$ identitou – kódovému slovu $\mathbf{a} \in \mathcal{K}$ priraďuje $\delta(\mathbf{a}) = \mathbf{a}$.

Ak bolo vyslané slovo $\mathbf{v}$ a došlo k chybe vyjadrenej slovom $\mathbf{e}$, prijmeme slovo $\mathbf{e} + \mathbf{v}$. Ak $\delta(\mathbf{e} + \mathbf{v}) = \mathbf{v}$, dekódovali sme správne.

Definícia 4.26. Hovoríme, že **lineárny kód $\mathcal{K}$ pri dekódovaní δ opravuje chybové slovo e, ak pre všetky $\mathbf{v} \in \mathcal{K}$ platí:**

$$\delta(\mathbf{e} + \mathbf{v}) = \mathbf{v} \ .$$

Definícia 4.27. Nech $\mathcal{K} \subseteq A^n$ je lineárny kód s kódovou abecedou A. Pre každé slovo $\mathbf{e} \in A^n$ definujeme

$$\mathbf{e} + \mathcal{K} = \{\mathbf{e} + \mathbf{v} \mid \mathbf{v} \in \mathcal{K}\} \ .$$

Množina $\mathbf{e} + \mathcal{K}$ sa volá **trieda slova e podľa kódu** $\mathcal{K}$.

Veta 4.26. *Nech $\mathcal{K} \subseteq A^n$ je lineárny (n, k)-kód s kódovou abecedou A, $|A| = p$. Pre ľubovoľné slová $\mathbf{e}$, $\mathbf{e}' \in A^n$ platí*

(i) Ak $\mathbf{e} - \mathbf{e}'$ je kódové slovo, potom $\mathbf{e} + \mathcal{K} = \mathbf{e}' + \mathcal{K}$.

(ii) Ak $\mathbf{e} - \mathbf{e}'$ nie je kódové slovo, potom $\mathbf{e} + \mathcal{K}$, $\mathbf{e}' + \mathcal{K}$ sú disjunktné.

(iii) Počet slov každej triedy sa rovná počtu kódových slov, t. j. $|\mathbf{e}+\mathcal{K}| = |\mathcal{K}| = p^k$ *a počet všetkých tried je* p^{n-k}.

Dôkaz. (i) Nech $(\mathbf{e}-\mathbf{e}') \in \mathcal{K}$, nech $\mathbf{v} \in \mathcal{K}$, a teda $(\mathbf{e}+\mathbf{v}) \in (\mathbf{e}+\mathcal{K})$. Položme $\mathbf{u} = \mathbf{v} + (\mathbf{e}-\mathbf{e}')$. Pretože $\mathcal{K}$ je lineárny priestor a $(\mathbf{e}-\mathbf{e}') \in \mathcal{K}$, je aj $\mathbf{u} \in \mathcal{K}$, a teda $(\mathbf{e}'+\mathbf{u}) \in (\mathbf{e}'+\mathcal{K})$. Ale $\mathbf{e}'+\mathbf{u} = \mathbf{e}'+\mathbf{v}+(\mathbf{e}-\mathbf{e}') = \mathbf{e}+\mathbf{v}$. Preto je $(\mathbf{e}+\mathbf{v}) \in (\mathbf{e}'+\mathcal{K})$. Ukázali sme, že $(\mathbf{e}+\mathcal{K}) \subseteq (\mathbf{e}'+\mathcal{K})$. Analogicky sa ukáže aj opačná inklúzia, a teda $(\mathbf{e}+\mathcal{K}) = (\mathbf{e}'+\mathcal{K})$.

(ii) Nech $(\mathbf{e}-\mathbf{e}') \notin \mathcal{K}$. Keby existovalo $\mathbf{w} \in (\mathbf{e}+\mathcal{K}) \cap (\mathbf{e}'+\mathcal{K})$, museli by existovať slová $\mathbf{v},\ \mathbf{v}' \in \mathcal{K}$ také, že

$$\mathbf{w} = \mathbf{e}+\mathbf{v}\ ,$$
$$\mathbf{w} = \mathbf{e}'+\mathbf{v}',$$

odkiaľ máme $\mathbf{e}+\mathbf{v} = \mathbf{e}'+\mathbf{v}'$ a ďalej $\mathbf{e}-\mathbf{e}' = \mathbf{v}'-\mathbf{v} \in \mathcal{K}$, lebo obe slová $\mathbf{v},\ \mathbf{v}'$ boli prvkami lineárneho priestoru $\mathcal{K}$. Z predpokladu, že $(\mathbf{e}+\mathcal{K}) \cap (\mathbf{e}'+\mathcal{K}) \neq \emptyset$ sme dostali $(\mathbf{e}-\mathbf{e}') \in \mathcal{K}$, čo je spor.

(iii) V úvahách bezprostredne po definícii 4.18 (str. 110) sme ukázali, že lineárny (n,k)-kód s p-prvkovou abecedou má p^k prvkov. Chceme ukázať, že $|\mathbf{e}+\mathcal{K}| = |\mathcal{K}| = p^k$. Na to stačí ukázať, že ak $\mathbf{u},\ \mathbf{w} \in \mathcal{K}$, $\mathbf{u} \neq \mathbf{b}$, potom $\mathbf{e}+\mathbf{u} \neq \mathbf{e}+\mathbf{w}$. Keby však $\mathbf{e}+\mathbf{u} = \mathbf{e}+\mathbf{w}$, potom by (po odčítaní $\mathbf{e}$ od oboch strán rovnice) $\mathbf{u} = \mathbf{w}$. Všetky triedy slov podľa kódu $\mathcal{K}$ majú rovnaký počet prvkov p^k. Keďže zjednotenie všetkých tried slov podľa kódu $\mathcal{K}$ je A^n a $|A^n| = p^k$ je sa počet všetkých rôznych tried podľa kódu $\mathcal{K}$ rovná číslu

$$\frac{|A^n|}{|\mathcal{K}|} = \frac{p^n}{p^k} = p^{n-k}.$$

■

Definícia 4.28. Štandardné dekódovanie. Definujeme úplné dekódovanie $\delta : A^n \to \mathcal{K}$ nasledovne: Z každej triedy podľa $\mathcal{K}$ vyberieme jedného reprezentanta triedy tak, aby jeho váha bola v danej triede minimálna. (Výber reprezentanta podľa kritéria minimálnej váhy nemusí byť jednoznačný – v tom prípade sa musíme rozhodnúť pre jedného s minimálnou váhou). Potom každé prijaté slovo $\mathbf{w} \in A^n$ dekódujeme ako $\mathbf{v} = \mathbf{w}-\mathbf{e}$, kde chybové slovo $\mathbf{e}$ je reprezentantom triedy slova $\mathbf{w}$, teda

$$\delta(\mathbf{w}) = \mathbf{w} - [\text{reprezentant triedy } (\mathbf{w}+\mathcal{K})].$$

Príklad 4.32. Binárny $(4,3)$-kód $\mathcal{K}$ celkovej parity má dve triedy

$$\begin{array}{lllllllll} 0000+\mathcal{K} & =\{0000 & 0011 & 0101 & 0110 & 1001 & 1010 & 1100 & 1111\} \\ 0001+\mathcal{K} & =\{0001 & 0010 & 0100 & 0111 & 1000 & 1011 & 1101 & 1110\} \end{array}$$

Trieda $0000+\mathcal{K}$ má jednoznačného reprezentanta – slovo 0000. Trieda $0001+\mathcal{K}$ môže mať za reprezentanta ľubovoľné zo slov 0001, 0010, 0100, 1000. Podľa toho, ktoré z týchto slov vyberieme za reprezentantov, štandardné dekódovanie opraví jednu chybu, ktorá vznikne na prvom, resp. druhom, treťom alebo štvrtom mieste. Ak vznikne chyba na inom mieste, štandardné dekódovanie nedekóduje správne. Pre nás to nie je prekvapujúce zistenie, lebo vieme, že kód celkovej parity má minimálnu vzdialenosť 2, a preto nemôže opravovať ani všetky jednoduché chyby.

Veta 4.27. *Štandardné dekódovanie δ opravuje práve tie chybové slová, ktoré sú reprezentantmi tried, t. j.*

$$\delta(\mathbf{v}+\mathbf{e})=\mathbf{v} \quad \textit{pre všetky } \mathbf{v}\in\mathcal{K}$$

práve vtedy, keď $\mathbf{e}$ je reprezentantom niektorej triedy podľa kódu $\mathcal{K}$.

Dôkaz. Ak je $\mathbf{e}$ reprezentantom svojej triedy a $\mathbf{v}\in\mathcal{K}$, potom slovo $\mathbf{v}+\mathbf{e}$ padne do triedy $\mathbf{e}+\mathcal{K}$ a dekóduje sa ako $\delta(\mathbf{e}+\mathbf{v})=\mathbf{e}+\mathbf{v}-\mathbf{e}=\mathbf{v}$ – podľa definície 4.26 dekódovanie δ opravuje chybové slovo $\mathbf{e}$.

Nech $\mathbf{e}'$ nie je reprezentantom svojej triedy, ktorá má za reprezentanta slovo $\mathbf{e}\neq\mathbf{e}'$. Platí $(\mathbf{e}-\mathbf{e}')\in\mathcal{K}$. Nech $\mathbf{v}\in\mathcal{K}$, potom slovo $\mathbf{v}+\mathbf{e}'$ padne do triedy $\mathbf{e}+\mathcal{K}$ a dekóduje sa ako $\delta(\mathbf{v}+\mathbf{e}')=\mathbf{v}+\mathbf{e}'-\mathbf{e}\neq\mathbf{v}$. Ak $\mathbf{e}'$ nie je reprezentantom svojej triedy, štandardné dekódovanie neopravuje slovo $\mathbf{e}'$.

Veta 4.28. *Štandardné dekódovanie δ je optimálne v tom zmysle, že neexistuje dekódovanie δ^*, ktoré by opravovalo tie isté chybové slová ako δ a navyše ešte niektoré ďalšie.*

Dôkaz. Vezmime $\mathbf{e}'\in(\mathbf{e}+\mathcal{K})$, nech $\mathbf{e}$ je reprezentantom triedy $\mathbf{e}+\mathcal{K}$, nech $\mathbf{e}\neq\mathbf{e}'$. Slovo $\mathbf{v}=\mathbf{e}'-\mathbf{e}$ je kódové a nenulové. Ak vyšleme slovo $\mathbf{v}$ a vznikne chyba pôsobením chybového slova $\mathbf{e}$, prijmeme slovo $\mathbf{v}+\mathbf{e}=\mathbf{e}'-\mathbf{e}+\mathbf{e}=\mathbf{e}'$. Keďže δ opravuje všetky slová, ktoré sú reprezentantami tried je $\delta(\mathbf{v}+\mathbf{e})=\delta(\mathbf{e}')=\mathbf{v}$. Keďže δ^* opravuje všetky slová, ktoré opravuje δ, je aj $\delta^*(\mathbf{e}')=\mathbf{v}$.

Môže dekódovanie δ^* opravovať slovo $\mathbf{e}'$? Keby áno, potom by muselo byť $\delta^*(\mathbf{o}+\mathbf{e}')=\mathbf{o}$, čo je v spore s tým, že $\delta^*(\mathbf{e}')=\mathbf{v}\neq\mathbf{o}$. ■

Veta 4.29. *Ak je $d = \Delta(\mathcal{K})$ minimálna vzdialenosť lineárneho kód $\mathcal{K}$, potom štandardné dekódovanie opraví všetky t-násobné chyby pre $t < \frac{d}{2}$.*

Dôkaz. Nech $\mathbf{e}$ je slovo váhy $t < \frac{d}{2}$. Nech $\mathbf{v} \in (\mathbf{e} + \mathcal{K})$, $\mathbf{v} \neq \mathbf{e}$, $\mathbf{v} = \mathbf{e} + \mathbf{u}$, $\mathbf{u} \in \mathcal{K}$. Je $\|\mathbf{u}\| \geq d$, $\|\mathbf{e}\| = t < \frac{d}{2}$. Preto počet nenulových znakov slova $\mathbf{v} = \mathbf{e} + \mathbf{u}$ je aspoň $d - t$ – t. j. $\|\mathbf{v}\| > d - t > t$. Preto je každé slovo $\mathbf{e}$ s Hammingovou váhou menšou než $\frac{d}{2}$ reprezentantom niektorej triedy slov podľa kódu $\mathcal{K}$. Keďže štandardné dekódovanie opravuje všetky chybové slová, ktoré sú reprezentantami tried, opravuje všetky chybové slová s Hammingovou váhou menšou než $\frac{d}{2}$, čo je ekvivalentné s tým, že štandardné dekódovanie opraví všetky t-násobné chyby. ■

Princípom štandardného dekódovania je určenie, v ktorej triede slov podľa kódu $\mathcal{K}$ sa dekódované slovo vyskytuje. Na to by príslušný dekódovací algoritmus musel prezrieť tzv. Slepianovu tabuľku všetkých slov dĺžky n abecedy A. Je to tabuľka, ktorá má toľko stĺpcov, koľko je tried podľa kódu $\mathcal{K}$ – p^{n-k} a toľko riadkov, koľko je kódových slov – p^k. V každom stĺpci tabuľky sú všetky slová jednej triedy, v prvom riadku tabuľky je reprezentant triedy. Po určení, v ktorom stĺpci sa dekódované slovo $\mathbf{w}$ nachádza dekódujeme tak, že od neho odčítame slovo v prvom riadku príslušného stĺpca.

	Trieda $\mathbf{e}_1 + \mathcal{K}$	Trieda $\mathbf{e}_2 + \mathcal{K}$		Trieda $\mathbf{e}_m + \mathcal{K}$
reprezentant	$\mathbf{e_1} = \mathbf{e_1} + \mathbf{o}$	$\mathbf{e_2} = \mathbf{e_2} + \mathbf{o}$	...	$\mathbf{e_m} = \mathbf{e_m} + \mathbf{o}$
prvky tried	$\mathbf{e_1} + \mathbf{u_1}$	$\mathbf{e_2} + \mathbf{u_1}$	...	$\mathbf{e_m} + \mathbf{u_1}$
	$\mathbf{e_1} + \mathbf{u_2}$	$\mathbf{e_2} + \mathbf{u_2}$	...	$\mathbf{e_m} + \mathbf{u_2}$
	...	...	...	...
	...	...	...	...
	...	...	...	...
	$\mathbf{e_1} + \mathbf{u_q}$	$\mathbf{e_2} + \mathbf{u_q}$	...	$\mathbf{e_m} + \mathbf{u_q}$

(4.54)

Slepianova tabuľka, $m = p^{n-k}$, $q = |\mathcal{K}| = p^k$.

Slepianova tabuľka má p^n prvkov, ktoré v najhoršom prípade musíme prehľadať všetky. Pre bežne používané binárne kódy dĺžky 64 by to znamenalo

v najhoršom prípade $2^{64} > 10^{19}$ prehľadaní. Šikovnou implementáciou možno úplné prehľadávanie nahradiť binárnym prehľadávaním, ktoré by v tomto prípade potrebovalo len 64 prístupov do tabuľky, ale nároky na príslušné údajové štruktúry ostávajú enormné.

Problém možno značne zredukovať, ak si uvedomíme, že všetky prvky triedy $\mathbf{e}+\mathcal{K}$ majú rovnaký syndróm ako jej reprezentant $\mathbf{e}$. Je to preto, lebo pre $\mathbf{v} \in \mathcal{K}$ a kontrolnú maticu $\mathbf{H}$ kódu $\mathcal{K}$ platí:

$$\mathbf{H}.(\mathbf{e}+\mathbf{v}) = \mathbf{H}.\mathbf{e} + \mathbf{H}.\mathbf{v} = \mathbf{H}.\mathbf{e} + \mathbf{o} = \mathbf{H}.\mathbf{e} \ .$$

Preto namiesto Slepianovej tabuľky stačí tabuľka s dvoma riadkami, kde v prvom riadku sú reprezentanti tried $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_m}$, $m = p^{n-k}$ a v druhom riadku sú príslušné syndrómy $\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_m}$.

reprezentant	$\mathbf{e}_1$	$\mathbf{e}_2$	...	$\mathbf{e}_m$
syndróm	$\mathbf{s}_1$	$\mathbf{s}_2$	...	$\mathbf{s}_m$

(4.55)

Teraz možno štandardný dekódovací algoritmus preformulovať nasledovne: Pre prijaté slovo $\mathbf{w}$ vypočítame jeho syndróm $\mathbf{s} = \mathbf{H}.\mathbf{w}$. V tabuľke (4.55) nájdeme reprezentanta $\mathbf{e}$ triedy s rovnakým syndrómom $\mathbf{s}$ a dekódujeme

$$\delta(\mathbf{w}) = \mathbf{w} - \mathbf{e} \ .$$

Tabuľka (4.55) má p^{n-k} stĺpcov a len dva riadky – jej rozsah je podstatne menší ako rozsah Slepianovej tabuľky. Navyše možno očakávať, že ani pri veľkej dĺžke n kódu $\mathcal{K}$ sa nebude číslo $n-k$ príliš zvyšovať, pretože znamená tiež počet kontrolných znakov kódu, a ten sa z hľadiska udržania dobrého informačného pomeru snažíme zbytočne nezvyšovať.

4.15 Hammingové kódy

Veta 4.30. *p-znakový lineárny kód opravuje jednoduché chyby práve vtedy, keď žiaden stĺpec jeho kontrolnej matice nie je skalárnym násobkom iného stĺpca. Špeciálne binárny lineárny kód opravuje jednoduché chyby práve vtedy, keď stĺpce jeho kontrolnej matice sú nenulové a navzájom rôzne.*

Dôkaz. Vieme, že kód $\mathcal{K}$ opravuje jednoduché chyby práve vtedy, keď $\Delta(\mathcal{K}) \geq 3$, čo podľa vety 4.24 (str. 120) nastáva práve vtedy, keď ľubovoľné dva stĺpce kontrolnej matice $\mathbf{H}$ sú lineárne nezávislé.

Vo všeobecnom prípade sú dva vektory $\mathbf{u}$, $\mathbf{v}$ lineárne nezávislé práve vtedy, keď jeden nie je skalárnym násobkom druhého, čo v prípade binárnej abecedy je práve vtedy, keď sú oba vektory $\mathbf{u}$, $\mathbf{v}$ nenulové a rôzne. ■

Definícia 4.29. Binárny lineárny (n,k)-kód sa nazýva **Hammingov kód**, ak jeho kontrolná matica $\mathbf{H}$ má za stĺpce všetky nenulové binárne slová dĺžky $n-k$, pričom každé z nich sa ako stĺpec matice $\mathbf{H}$ vyskytuje práve raz.

Ak sa majú v matici $\mathbf{H}$ všetky nenulové binárne slová dĺžky $n-k$ vyskytovať práve raz, musí sa počet stĺpcov v tejto matici rovnať

$$n = 2^{(n-k)} - 1.$$

Preto môže existovať len Hammingov (n,k)-kód pre

$$(n,k) = (3,1),\ (7,4),\ (15,11),\ (31,26),\ldots(2^m-1, 2^m-m-1),\ldots\ .$$

Všimnime si ešte, že informačný pomer (4.39) (str. 105) s rastúcim m rýchlo rastie k 1. Napr. pre $m = 6$ Hammingov $(63,57)$-kód má informačný pomer $\dfrac{57}{63} > 0.9$.

Definícia 4.30. Dekódovanie Hammingovho kódu. Predpokladajme, že stĺpce kontrolnej matice $\mathbf{H}$ sú usporiadané tak, že tvoria binárne rozvoje čísel $1, 2, \ldots, 2^{m-1}$. Prijmeme vektor $\mathbf{w}$ a vypočítame jeho syndróm $\mathbf{s} = \mathbf{Hw}$. Ak $\mathbf{s} = \mathbf{o}$, slovo $\mathbf{w}$ nemeníme. Ak $\mathbf{s} \neq \mathbf{o}$, slovo $\mathbf{s}$ je binárnym rozvojom čísla i a my zmeníme i-ty znak prijatého slova $\mathbf{w}$. Presnejšie

$$\delta(\mathbf{w}) = \begin{cases} \mathbf{w}, & \text{ak } \mathbf{s} = \mathbf{o} \\ \mathbf{w} - \mathbf{e_i}, & \text{ak } \mathbf{s} \text{ je binárnym rozvojom čísla } i, \end{cases} \tag{4.56}$$

kde $\mathbf{e}_i$ je slovo s jednotkou na mieste i.

Veta 4.31. *Dekódovanie δ definované v* (4.56) *opravuje jednoduché chyby. Presnejšie: Ak sa slovo* $\mathbf{w}$ *líši od niektorého kódového slova* $\mathbf{v}$ *nanajvýš v jednom znaku, potom* $\delta(\mathbf{w}) = \mathbf{v}$.

Dôkaz. Ak $\mathbf{w} = \mathbf{v}$, potom aj $\mathbf{w}$ je kódové slovo a platí $\mathbf{Hw} = \mathbf{Hv} = \mathbf{o}$, a v tom prípade $\delta(\mathbf{w}) = \mathbf{w} = \mathbf{v}$.

Nech sa slová $\mathbf{v}$, $\mathbf{w}$ líšia práve v jednom znaku, t. j. $\mathbf{w} = \mathbf{v} + \mathbf{e_i}$, kde $\mathbf{e}_i$ je slovo s jednotkou na mieste i, $i \in \{1, 2, \ldots, n\}$. Potom

$$\mathbf{Hw} = \mathbf{H}(\mathbf{v} + \mathbf{e_i}) = \mathbf{Hv} + \mathbf{He_i} = \mathbf{He_i}\ .$$

Ale $\mathbf{He_i}$ je i-ty stĺpec matice $\mathbf{H}$ a ten je rozvojom čísla i. Ak budeme dekódovať predpisom $\delta(\mathbf{w}) = \mathbf{w} - \mathbf{e_i} = \mathbf{v}$, budeme dekódovať správne. ■

Medzi kódmi, ktoré opravujú t chýb, sú najekonomickejšie tzv. perfektné kódy. Podľa definície 4.13 (str. 101) blokový kód $\mathcal{K}$ dĺžky n je t-perfektný, ak množina gúľ $\{G_t(\mathbf{a}) \mid \mathbf{a} \in \mathcal{K}\}$ tvorí rozklad množiny A^n všetkých slov dĺžky n.

Veta 4.32. *Lineárny kód je t-perfektný práve vtedy, keď množina všetkých slov váhy menšej než alebo rovnajúcej sa číslu t tvorí systém všetkých reprezentantov všetkých tried slov podľa kódu $\mathcal{K}$.*

Dôkaz. Prv, než začneme dokazovať tvrdenie vety, si všimneme, že ľubovoľné slovo $\mathbf{a} \in A^n$ môže byť reprezentantom niektorej triedy kódu $\mathcal{K}$ – totiž triedy $\mathbf{a}+\mathcal{K}$. Na to, aby sme dokázali, že množina všetkých slov váhy menšej než alebo rovnajúcej sa číslu t tvorí systém všetkých reprezentantov všetkých tried slov podľa kódu $\mathcal{K}$ stačí ukázať dve skutočnosti, a to že

- každá trieda má reprezentanta s váhou menšou než alebo rovnajúcou sa číslu t
- ak $\mathbf{e_1}$, $\mathbf{e_2}$ sú dve slová také, že $\|\mathbf{e}_1\| \leq t$, $\|\mathbf{e}_2\| \leq t$, potom $\mathbf{e}_1 + \mathcal{K}$, $\mathbf{e}_1 + \mathcal{K}$ sú dve rôzne triedy, t. j. $\mathbf{e}_2 \notin (\mathbf{e}_1 + \mathcal{K})$

1. Nech $\mathcal{K}$ je t-perfektný lineárny kód – t. j. pre každé slovo $\mathbf{a} \in A^n$ existuje práve jedno kódové slovo $\mathbf{b} \in \mathcal{K}$ také, že vzdialenosť slov $\mathbf{a}$, $\mathbf{b}$ je menšia alebo rovnajúca sa t, t. j. $d(\mathbf{a}, \mathbf{b}) \leq t$. Označme $\mathbf{e} = \mathbf{a} - \mathbf{b}$. Pretože Hammingova vzdialenosť slov $\mathbf{a}$, $\mathbf{b}$ je menšia alebo rovnajúca sa t, je $\|\mathbf{e}\| \leq t$. Potom je $\mathbf{a} = \mathbf{e} + \mathbf{b}$. Každá trieda $\mathbf{a} + \mathcal{K}$ má reprezentanta $\mathbf{e}$ s váhou menšou alebo rovnajúcou sa t.
Keby existovali dve slová $\mathbf{e}_1$, $\mathbf{e}_2$ také, že $\|\mathbf{e}_1\| \leq t$, $\|\mathbf{e}_2\| \leq t$ a $\mathbf{e}_2 \in (\mathbf{e}_1 + \mathcal{K})$, potom $\mathbf{e_2} - \mathbf{e_1} \in \mathcal{K}$ a $\|\mathbf{e_2} - \mathbf{e_1}\| \leq 2t$. Z poslednej nerovnosti vyplýva pre minimálnu vzdialenosť $\Delta(\mathcal{K})$ kódu $\mathcal{K}$: $\Delta(\mathcal{K}) \leq 2t$, čo je v spore s predpokladom, že $\mathcal{K}$ opravuje t chýb. Podľa vety 4.13 (str. 101) totiž kód $\mathcal{K}$ opravuje t chýb práve vtedy, keď $\Delta(\mathcal{K}) \geq 2t + 1$.

2. Nech množina všetkých slov váhy menšej alebo rovnej než t tvorí systém všetkých reprezentantov všetkých tried slov podľa kódu $\mathcal{K}$. Najprv ukážeme, že $\Delta(\mathcal{K}) \geq 2t+1$. Keby totiž existovalo $\mathbf{a} \in \mathcal{K}$ také, že $\|\mathbf{a}\| \leq 2t+1$, bolo by možné vyjadriť $\mathbf{a} = \mathbf{e_1} - \mathbf{e_2}$, kde $\|\mathbf{e}_1\| \leq t$, $\|\mathbf{e}_2\| \leq t$ a $\mathbf{e_1} \neq \mathbf{e_2}$. Podľa (i) vety 4.26 (str. 122) by potom $(\mathbf{e}_1 + \mathcal{K}) = (\mathbf{e}_2 + \mathcal{K})$, čo by bolo v spore s predpokladom, že $\mathbf{e_1}$, $\mathbf{e_2}$ sú reprezentantmi rôznych tried. Ak je teda $\Delta(\mathcal{K}) \geq 2t + 1$, všetky gule $\{G_t(\mathbf{a}) \mid \mathbf{a} \in \mathcal{K}\}$ sú po dvoch disjunktné.
Teraz ukážeme, že pre každé $\mathbf{a} \in A^n$ existuje guľa $G_t(\mathbf{b})$, $\mathbf{b} \in \mathcal{K}$ taká, že $\mathbf{a} \in G_t(\mathbf{b})$. Podľa predpokladu existuje $\mathbf{e} \in A^n$, $\|\mathbf{e}\| \leq t$ také, že $\mathbf{a} \in (\mathbf{e} + \mathcal{K})$. Dá sa teda písať $\mathbf{a} = \mathbf{e} + \mathbf{b}$ pre nejaké $\mathbf{b} \in \mathcal{K}$. Odtiaľ $\mathbf{a} - \mathbf{b} = \mathbf{e}$, a preto $d(\mathbf{a}, \mathbf{b}) = \|(\mathbf{a} - \mathbf{b})\| = \|\mathbf{e}\| \leq t$ a teda $\mathbf{a} \in G_t(\mathbf{b})$. Systém gúľ $\{G_t(\mathbf{a}) \mid \mathbf{a} \in \mathcal{K}\}$ tvorí rozklad množiny A^n, a preto je kód $\mathcal{K}$ t-perfektný. ■

Veta 4.33. *Hammingové binárne kódy sú* 1*-perfektné. Každý* 1*-perfektný binárny lineárny kód je Hammingov.*

Dôkaz. Hammingov kód dĺžky $2^m - 1$ má m kontrolných znakov a podľa tvrdenia (iii) vety 4.26 (str. 122) má 2^m tried. Označme $\mathbf{e}_0 = \mathbf{o}$ – nulové slovo dĺžky $2^m - 1$. Ďalej označme pre $i = 1, 2, \ldots, 2^m - 1$

$$\mathbf{e}_i = \left[\ 0 \quad 0 \quad \ldots \quad 0 \quad 1 \quad 0 \ldots \quad 0\ \right],$$

t. j. vektor $\mathbf{e}_i$ má všade nuly okrem miesta i, na ktorom má znak 1. Všetky $\mathbf{e}_i$ pre $i = 1, 2, \ldots, 2^m - 1$ sú nekódové slová.

Skúmajme triedy $\mathbf{e}_i + \mathcal{K}$ pre $i = 0, 1, 2, \ldots, 2^m - 1$. Trieda $\mathbf{e}_0 + \mathcal{K}$ je totožná s množinou kódových slov $\mathcal{K}$, a je preto rôzna od ostatných tried. Keby boli dve triedy $\mathbf{e}_i + \mathcal{K}$, $\mathbf{e}_j + \mathcal{K}$ totožné pre $i \neq j$, potom by $\mathbf{e_i} - \mathbf{e_j} \in \mathcal{K}$, čo by znamenalo lineárnu závislosť i-teho a j-teho stĺpca kontrolnej matice kódu $\mathcal{K}$, (čo je v prípade binárneho kódu rovnosť príslušných stĺpcov). Hammingov kód má však kontrolnú maticu, v ktorej žiadne dva stĺpce nie sú rovnaké.

Pretože Hammingov kód $\mathcal{K}$ má 2^m tried a my sme ukázali, že všetky triedy typu $\mathbf{e}_i + \mathcal{K}$ pre $i = 0, 1, 2, \ldots, 2^m - 1$ sú rôzne (a je ich 2^m), nemôže existovať žiadna ďalšia trieda. Množina všetkých slov dĺžky ≤ 1 tvorí systém reprezentantov všetkých tried Hammingovho kódu $\mathcal{K}$, a preto je tento kód 1-perfektný.

Majme binárny lineárny kód $\mathcal{K}$ s m kontrolnými znakmi, ktorý je 1 perfektný. Podľa tvrdenia (iii) vety 4.26 (str. 122) kód $\mathcal{K}$ má 2^m tried. Nech má tento kód kontrolnú maticu $\mathbf{H}$ typu $n \times m$. Podľa vety 4.30 musia byť všetky stĺpce matice $\mathbf{H}$ nenulové a rôzne. Preto pre počet stĺpcov matice $\mathbf{H}$ platí $n \leq 2^m - 1$. Pretože je kód $\mathcal{K}$ perfektný, podľa vety 4.32 (str. 128) sú všetky binárne slová dĺžky n s váhou nula alebo jedna práve všetci reprezentanti tried. Takýchto slov je $n+1$ (nulové slovo a všetky slová typu $\mathbf{e}_i$ s práve jednou jednotkou na i-tom mieste). Je preto

$$n + 1 = 2^m,$$

čiže

$$n = 2^m - 1.$$

Kontrolná matica kódu $\mathcal{K}$ je matica typu $(2^m - 1) \times m$ a jej stĺpce sú práve všetky rôzne binárne nenulové slová dĺžky m. $\mathcal{K}$ je teda Hammingovým kódom. ■

Definícia 4.31. Rozšírený Hammingov binárny kód je binárny kód, ktorý vznikne rozšírením Hammingovho kódu o znak celkovej kontroly parity.

Rozšírený Hammingov kód je $(2^m, 2^m - m - 1)$-kód všetkých slov $\mathbf{v} = v_1 v_2 \dots v_{2^m}$ takých, že $v_1 v_2 \dots v_{2^m-1}$ je kódové slovo Hammingovho kódu a $v_1 + v_2 + \cdots + v_{2^m} = 0$. Jeho minimálna váha je 4. Tento kód opravuje jednoduché chyby a objavuje trojnásobné chyby.

Poznámka. Veta 4.30 dáva návod, ako definovať p-znakový Hammingov kód. Je to kód s kontrolnou maticou $\mathbf{H}$ takou, že

(i) žiaden stĺpec nie je skalárnym násobkom iného stĺpca

(ii) každé nenulové slovo je skalárnym násobkom niektorého stĺpca matice $\mathbf{H}$.

Maticu $\mathbf{H}$ môžeme zostaviť napríklad zo všetkých stĺpcov rovnakej dĺžky takých, ktoré majú prvý nenulový znak 1. Dá sa ukázať, že p-znakové Hammingové kódy majú mnohé vlastnosti rovnaké resp. analogické ako binárne Hammingové kódy. Tak napríklad všetky Hammingové kódy sú 1-perfektné.

4.16 Golayov kód*

Označme $\mathbf{B}$ štvorcovú maticu typu 11×11, ktorej prvý riadok obsahuje binárne slovo 11011100010 a ostatné riadky vzniknú pravými rotáciami prvého riadku, t. j.

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & & 1 & 1 & 1 & & & & 1 & \\ & 1 & 1 & & 1 & 1 & 1 & & & & 1 \\ 1 & & 1 & 1 & & 1 & 1 & 1 & & & \\ & 1 & & 1 & 1 & & 1 & 1 & 1 & & \\ & & 1 & & 1 & 1 & & 1 & 1 & 1 & \\ & & & 1 & & 1 & 1 & & 1 & 1 & 1 \\ 1 & & & & 1 & & 1 & 1 & & 1 & 1 \\ 1 & 1 & & & & 1 & & 1 & 1 & & 1 \\ 1 & 1 & 1 & & & & 1 & & 1 & 1 & \\ & 1 & 1 & 1 & & & & 1 & & 1 & 1 \\ 1 & & 1 & 1 & 1 & & & & 1 & & 1 \end{bmatrix}. \tag{4.57}$$

Binárne slovo 11011100010 má na mieste i jednotku práve vtedy, keď je $i-1$ štvorcom modulo 11, t. j. ak $i-1 = 0^2$, 1^2, 2^2, 3^2, $4^2 \equiv 5$ a $5^2 \equiv 3$. V celej časti 4.16 budeme predpokladať, že matica $\mathbf{B}$ je daná vzťahom (4.57).

Definícia 4.32. Golayov kód G_{23} je systematický binárny kód dĺžky 23 s generujúcou maticou $\mathbf{G}_{23}$ definovanou

$$\mathbf{G}_{23} = \left[\begin{array}{c|c} \mathbf{E}_{12\times 12} & \begin{array}{c}\mathbf{B}_{11\times 11}\\ \hline 11\ldots 11\end{array} \end{array}\right],$$

kde $\mathbf{E}_{12\times 12}$ je jednotková matica typu 12×12, $\mathbf{B}_{11\times 11}$ je štvorcová matica typu 11×11 definovaná v (4.57).

Golayov kód G_{24} je systematický binárny kód dĺžky 24 s generujúcou maticou $\mathbf{G}_{24}$, ktorá vznikne z matice $\mathbf{G}_{23}$ pridaním stĺpca $11\ldots 10$, t. j.

$$\mathbf{G}_{24} = \left[\begin{array}{c|c|c} \mathbf{E}_{12\times 12} & \begin{array}{c}\mathbf{B}_{11\times 11}\\ \hline 11\ldots 11\end{array} & \begin{array}{c}1\\1\\\ldots\\1\\ \hline 0\end{array} \end{array}\right]$$

$$\left[\begin{array}{cccccccccccccccccccccccc}
1 & & & & & & & & & & & 1 & 1 & & 1 & 1 & 1 & & & 1 & & 1 \\
 & 1 & & & & & & & & & & & 1 & 1 & & 1 & 1 & 1 & & & & 1 & 1 \\
 & & 1 & & & & & & & & & 1 & & 1 & 1 & & 1 & 1 & 1 & & & & 1 \\
 & & & 1 & & & & & & & & & 1 & & 1 & 1 & & 1 & 1 & 1 & & & 1 \\
 & & & & 1 & & & & & & & & & 1 & & 1 & 1 & & 1 & 1 & 1 & & 1 \\
 & & & & & 1 & & & & & & & & & 1 & & 1 & 1 & & 1 & 1 & 1 & 1 \\
 & & & & & & 1 & & & & & 1 & & & & 1 & & 1 & 1 & & 1 & 1 & 1 \\
 & & & & & & & 1 & & & & 1 & 1 & & & & 1 & & 1 & 1 & & 1 & 1 \\
 & & & & & & & & 1 & & & 1 & 1 & 1 & & & & 1 & & 1 & 1 & & 1 \\
 & & & & & & & & & 1 & & & 1 & 1 & 1 & & & & 1 & & 1 & 1 & 1 \\
 & & & & & & & & & & 1 & 1 & & 1 & 1 & 1 & & & & 1 & & 1 & 1 \\
 & & & & & & & & & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}\right]$$

Generujúca matica Golayovho kódu $\mathbf{G}_{24}$.

Vlastnosti kódov G_{24}**,** G_{23}**.**

- Počet informačných znakov kódu G_{24} je 12, počet kontrolných znakov je tiež 12.

- Kód G_{24} je samoduálny – jeho kontrolná matica je aj jeho generujúcou maticou (na to stačí overiť, že skalárny súčin ľubovoľných dvoch riadkov matice $\mathbf{G}_{24}$ sa rovná 0).
- Minimálna vzdialenosť kódu G_{24} je 8
- Kód G_{23} je $(23, 12)$-kód, ktorý je 3-perfektný.

Veta 4.34. Tietäväinen, Van Lint. *Jediné netriviálne perfektné binárne kódy sú tieto:*

a) Hammingove kódy pre jednoduché chyby,

b) Golayov kód G_{23} *pre trojnásobné chyby a kódy s ním ekvivalentné,*

c) opakovacie kódy dĺžky $2t + 1$ *pre* t*-násobné chyby, kde* $t = 1, 2, 3, \ldots$.

Dôkaz vlastností Golayových kódov i poslednej vety presahujú rámec tejto publikácie, preto ich neuvádzame. Čitateľ ich nájde v knihe [1].

K zaujímavým ternárnym kódom patrí perfektný Golayov ternárny $(11, 6)$-kód opravujúci trojnásobné chyby. Jeho generujúca matica je v tvare

$$\mathbf{G}_{11} = \left[\begin{array}{c|c} \mathbf{E}_{6\times 6} & \begin{array}{c} \mathbf{D}_{5\times 5} \\ \hline 11\ldots11 \end{array} \end{array}\right],$$

kde $\mathbf{E}_{6\times 6}$ je jednotková matica typu 6×6 a kde $\mathbf{D}_{5\times 5}$ je matica, ktorej riadky tvoria všetky cyklické pravé rotácie slova 01221. Okrem tohoto kódu (a kódov s ním ekvivalentných) sú jediné perfektné ternárne netriviálne kódy Hammingove a opakovacie kódy dĺžky $2t + 1$.

V prípade abecedy s viac ako troma znakmi sú jediné perfektné netriviálne kódy Hammingove a opakovacie kódy dĺžky $2t + 1$.

Na záver kapitoly o kódovaní treba povedať, že jej náplň tvorí iba úvod do teórie a praxe kódovania. Množstvo metód a kódov sa do tejto kapitoly nevošlo nielen z priestorových dôvodov, ale aj preto, lebo sú založené na pojmoch konečnej algebry ako boolovský polynóm, okruhy polynómov, konečné telesá atď. Takými sú Reedove-Mullerove kódy, cyklické kódy, BCH kódy atď. Čitateľ však túto problematiku môže nájsť v použitej literatúre [1], [2], [11]. Znalosť tejto kapitoly mu značne pomôže orientovať sa v problematike kódovania.

Kapitola 5

Prenosové kanály a ich kapacita

5.1 Bezporuchové kanály

Prenosový kanál si môžeme modelovať ako komunikačné zariadenie so vstupom a výstupom. Vstup dokáže spracovávať znaky vstupnej abecedy Y, z výstupu kanála vystupujú znaky výstupnej abecedy Z. Vo väčšine prípadov $Y = Z$, ale to nemusí byť pravidlom, preto budeme vstupnú a výstupnú abecedu odlišovať označením.

Príklad 5.1. Nech vstupnou abecedou Y skúmaného kanála je abeceda $Y = \{0, 1\}$ reprezentovaná napäťovými úrovňami $0 = L$-(low – nízka – napr. 0.7 V) a $1 = H$-(high – vysoká – napr. 5.5 V). Ak sa na výstupe objaví úroveň 3.1 V, nevieme rozhodnúť, či sme prijali nulu alebo jednotku. Preto na výstupe úrovne z intervalu $\langle 0.7, 2.3\rangle$ budeme interpretovať ako 0, úrovne z intervalu $\langle 3.9, 5.5\rangle$ ako 1 a úrovne z intervalu $(2.3, 3, 9)$ ako chybne prijatý znak $*$. Výstupnou abecedou bude v tomto prípade $Z = \{0, 1, *\}$.

Príklad 5.2. Nech vstupnú abecedu Y kanála tvorí množina všetkých 8-bitových čísel s párnou paritou. Ak ide o poruchový kanál, môžu sa na výstupe objaviť aj 8-bitové čísla s nepárnou paritou. Výstupnou abecedou Z kanála je množina všetkých 8-bitových čísel.

Do vstupu kanála prichádza v diskrétnych časových okamihoch $i = 1, 2, 3, \ldots$ postupnosť znakov $y_1, y_2, y_3, \ldots$ a v odpovedajúcich časových okamihoch sa

na výstupe objavuje postupnosť $z_1, z_2, z_3, \ldots$, t. j. ak sa v okamihu i objaví na vstupe znak y_i, potom v jemu prislúchajúcemu okamihu sa na výstupe objaví znak z_i. (V istom priblížení môžeme predpokladať, že v okamihu vstupu znaku y_i sa na výstupe kanála objaví znak z_i. Tento predpoklad síce odporuje fyzikálnym zákonom, podľa ktorých sa aj najrýchlejšie možné nosiče informácie – fotóny pohybujú konečnou rýchlosťou, avšak vo väčšine prípadov je oneskorenie výstupu za vstupom z nášho hľadiska zanedbateľné.)

Jednoduchším prípadom prenosového kanála je **bezporuchový kanál**, pri ktorom v čase i prijatý znak z_i závisí len na vyslanom znaku y_i – t. j.

$$z_i = f_i(y_i) \ ,$$

v tom prípade hovoríme, že sa jedná o **kanál bez pamäte**[1] , resp. v čase i prijatý znak z_i jednoznačne závisí len od vyslaného slova $y_1, y_2, \ldots, y_i$ – t. j.

$$z_i = F_i(y_1, y_2, \ldots, y_i) \ ,$$

v tom prípade hovoríme, že ide o **prenosový kanál s pamäťou**.[2]

Študujú sa aj **kanály s konečnou pamäťou**, kedy výstupný znak z_i závisí iba od posledných m vyslaných znakov, t. j.

$$z_i = F_i(y_{i-m+1}, y_{i-m+2}, \ldots, y_i) \ .$$

Od kanálov, ktorými sa budeme zaoberať, budeme tiež žiadať jednu samozrejmú vlastnosť – výstupný znak z_i nesmie nijako závisieť na žiadnom vstupnom znaku y_{i+k}, $k > 0$. To, čo sa na výstupe objaví v čase i, závisí len na vstupoch $y_1, y_2, \ldots, y_i$, ale nesmie nijako závisieť na vstupoch po čase i. Túto vlastnosť nazývame **nepredvídavosť**. Bezporuchový kanál teda jednoznačne popíšeme súborom funkcií $\{f_i\}_{i=1,2,\ldots}$ resp. $\{F_i\}_{i=1,2,\ldots}$.

5.2 Prenosové kanály so šumom

V reálnych situáciách je bezporuchový kanál skôr výnimkou, ako pravidlom. Priemyselné rušenie, počasie, atmosferické výboje, statická elektrina, kozmický

[1] Najbežnejší prípad je ten, kedy $Y = Z$ a f_i je identita na $Y = Z$ pre každé i. Vo všeobecnejších prípadoch môže funkcia f_i závisieť aj od časového okamihu i.

[2] Napríklad ⟨`CapsLock`⟩ spôsobí že po jeho zadaní klávesnica počítača vysiela po stlačení alfabetických kláves veľké písmena, po jeho opätovnom stlačení sa prepne do módu malých písmen. Tento kanál si „pamätá“ históriu vstupu a podľa toho generuje výstupné znaky.

Podobne po vstupe ⟨`Alt`⟩/⟨`Shift`⟩ pod OS Windows sa klávesnica prepne na diakritiku.

šum, vtáky v okolí vysielacích a prijímacích antén a mnohé iné vplyvy spôsobujú poruchy pri prenosoch.[3] Ak vyšleme cez rušený kanál vstupné slovo $y_1, y_2, \ldots, y_i$, môžeme vplyvom porúch prijať ľubovoľné výstupné slovo $z_1, z_2, \ldots, z_i$, pravda, s rôznou pravdepodobnosťou. Podmienenú pravdepodobnosť prijatia slova $z_1, z_2, \ldots, z_i$ za predpokladu, že bolo vyslané slovo $y_1, y_2, \ldots, y_i$, označíme symbolom $\nu(z_1, z_2, \ldots, z_i|y_1, y_2, \ldots, y_i)$. Keďže vstupná abeceda Y, výstupná abeceda Z a funkcia $\nu : \bigcup_{i=1}^{\infty}(Z^i \times Y^i) \to \langle 0, 1\rangle$ plne charakterizujú prenosový kanál, môžeme definovať

Definícia 5.1. Prenosový kanál $\mathcal{C}$ je usporiadaná trojica $\mathcal{C} = (Y, Z, \nu)$, kde Y je vstupná abeceda, Z je výstupná abeceda a $\nu : \bigcup_{i=1}^{\infty}(Z^i \times Y^i) \to \langle 0, 1\rangle$, pričom $\nu(z_1, z_2, \ldots, z_i|y_1, y_2, \ldots, y_i)$ je podmienená pravdepodobnosť, že na výstupe sa objaví slovo $z_1, z_2, \ldots, z_i$ za predpokladu, že na vstupe bolo slovo $y_1, y_2, \ldots, y_i$.

Označme $\nu_i(z_i|y_1, y_2, \ldots, y_i)$ podmienenú pravdepodobnosť javu, že sa v čase i objaví na výstupe znak z_i za predpokladu, že na vstupe bolo slovo $y_1, y_2, \ldots, y_i$. Potom

$$\nu_i(z_i|y_1, y_2, \ldots, y_i) = \sum_{z_1, z_2, \ldots, z_{i-1}} \nu(z_1, z_2, \ldots, z_i|y_1, y_2, \ldots, y_i).$$

Hovoríme, že kanál $\mathcal{C}$ je **kanál bez pamäte**, ak $\nu_i(z_i|y_1, y_2, \ldots, y_i)$ závisí iba na y_i, t. j. ak

$$\nu_i(z_i|y_1, y_2, \ldots, y_i) = \nu_i(z_i|y_i).$$

Ak navyše $\nu_i(z_i|y_i)$ nezávisí na i, t. j. ak $\nu_i(z_i|y_i) = \nu(z_i|y_i)$, hovoríme, že $\mathcal{C}$ je **stacionárny kanál bez pamäte**.
Ak

$$\nu(z_1, z_2, \ldots, z_i|y_1, y_2, \ldots, y_i) = \nu(z_1|y_1)\nu(z_2|y_2)\ldots\nu(z_i|y_i) = \prod_{k=1}^{i} \nu(z_k|y_k),$$

hovoríme o **stacionárnom nezávislom kanáli**.

[3] Niekedy je kanálom aj človek, napr. číta (očami) čísla (tovarov, bankových účtov, železničných vozňov) a prenáša ich tak, že ich cez klávesnicu vkladá do registračnej pokladne, alebo do počítača, alebo ručne zapisuje do zoznamu. Z omylnosti človeka vyplýva, že tento kanál nie je bezporuchový – je to kanál so šumom. V prenosových kanáloch so šumom preto používame kódy, ktoré objavujú, alebo opravujú dovolený počet chýb.

5.3 Stacionárny nezávislý kanál

Majme stacionárny nezávislý kanál so vstupnou abecedou $A = \{a_1, a_2, \ldots, a_n\}$ a výstupnou abecedou $B = \{b_1, b_2, \ldots, b_r\}$. Označme $q_{ij} = \nu(b_j|a_i)$ pravdepodobnosť, že ak je na vstupe kanála vstupný znak a_i, na výstupe sa objaví znak b_j.
Hodnoty q_{ij} sa volajú **prenosové pravdepodobnosti** a matica typu $n \times r$

$$\mathbf{Q} = \begin{pmatrix} q_{11} & q_{12} & \ldots & q_{1r} \\ q_{21} & q_{22} & \ldots & q_{2r} \\ \ldots & \ldots & \ldots & \ldots \\ q_{n1} & q_{n2} & \ldots & q_{nr} \end{pmatrix}$$

je **matica prenosových pravdepodobností**. Poznamenajme, že súčet prvkov každého riadku matice $\mathbf{Q}$ sa rovná 1, t. j. $\sum_{j=1}^{r} q_{kj} = 1$ pre každé $k = 1, 2, \ldots, n$.

Nech $p_i = P(a_i)$ je pravdepodobnosť javu, že sa na vstupe kanála objaví znak a_i. Pravdepodobnosť javu $P(a_i \cap b_j)$, že na vstupe kanála bude znak a_i a na jeho výstupe znak b_j, vypočítame ako

$$P(a_i \cap b_j) = p_i q_{ij}.$$

Pravdepodobnosť $P(b_j)$, že sa na výstupe kanála objaví b_j vypočítame ako súčet pravdepodobností $P(a_1 \cap b_j) + P(a_2 \cap b_j) + \cdots + P(a_n \cap b_j)$, t. j.

$$P(b_j) = \sum_{t=1}^{n} p_t q_{tj}.$$

Na výskyt znaku a_i na vstupe kanála resp. znaku b_j na výstupe kanála sa môžeme pozerať ako na výsledky pokusov

$$\begin{aligned} \mathbf{A} &= \big\{\{a_1\}, \{a_2\}, \ldots, \{a_n\}\big\}, \\ \mathbf{B} &= \big\{\{b_1\}, \{b_2\}, \ldots, \{b_r\}\big\}. \end{aligned}$$

Príjemcu správ na výstupe kanála zaujíma, aký znak bol vyslaný – teda aký bol výsledok pokusu $\mathbf{A}$. Má však k dispozícii len výsledok pokusu $\mathbf{B}$. V časti 2.7 sme ukázali, že stredná hodnota informácie obsiahnutá v pokuse $\mathbf{B}$ o pokuse $\mathbf{A}$ sa dá vyjadriť ako spoločná informácia $I(\mathbf{A}, \mathbf{B})$ pokusov $\mathbf{A}$, $\mathbf{B}$, pre ktorú využijeme vzťah (2.14) z vety 2.14 (str. 46)

$$I(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i \cap B_j). \log_2 \left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)} \right). \tag{5.1}$$

Vzťah (5.1) môžeme prepísať pomocou pravdepodobností p_i, q_{ij} nasledovne:

$$\begin{aligned} I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{r} P(a_i \cap b_j)\log_2 \frac{P(a_i \cap b_j)}{P(a_i)P(b_j)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{r} p_i q_{ij}\log_2 \frac{p_i q_{ij}}{p_i \sum_{t=1}^{n} p_t q_{tj}} \\ &= \sum_{i=1}^{n} p_i \sum_{j=1}^{r} q_{ij}\log_2 \frac{q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}}. \end{aligned} \tag{5.2}$$

Ak sa bude pokus **A** mnohokrát nezávisle opakovať (t. j. na vstupe kanála sa budú objavovať výstupy zo stacionárneho nezávislého zdroja s abecedou A a pravdepodobnosťami p_i, $i = 1, 2, \ldots, n$), výraz (5.1), resp. (5.2) je stredné množstvo informácie prenesené kanálom pripadajúce na jeden znak.

Špeciálnym prípadom diskrétneho kanála bez pamäte je symetrický binárny kanál, ktorého vstupná abeceda je $A = \{0, 1\}$, výstupná abeceda je $B = \{0, 1\}$ a matica prenosových pravdepodobností je

$$\mathbf{Q} = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix}, \tag{5.3}$$

kde $0 \leq q \leq 1$. V tomto prípade $n = 2$ a tiež $r = 2$.
Všimnime si, že pre $q = 1/2$ je

$$\mathbf{Q} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \tag{5.4}$$

a preto

$$\begin{aligned} I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{2} p_i \sum_{j=1}^{2} \frac{1}{2}\log_2 \frac{1/2}{\sum_{t=1}^{2} p_t . 1/2} \\ &= \sum_{i=1}^{2} p_i \sum_{j=1}^{2} \frac{1}{2}\log_2 \frac{1/2}{(1/2).\sum_{t=1}^{2} p_t} \\ &= \sum_{i=1}^{2} p_i \sum_{j=1}^{2} \frac{1}{2}\log_2 1 = 0 \end{aligned}$$

pre akékoľvek hodnoty pravdepodobností p_1, p_2. Kanál v tomto prípade neprenáša žiadnu informáciu.

Hľadajme pravdepodobnosti $p_1, p_2, \ldots, p_n$, pre ktoré je množstvo prenesenej informácie na jeden znak maximálne, t. j. hľadáme viazaný extrém funkcie (5.2) pri podmienke $\sum_{i=1}^{n} p_i = 1$ a pri podmienke $p_i \geq 0$ pre $i = 1, 2, \ldots, n$. Na to môžeme použiť metódu Lagrangeových multiplikátorov.
Položme

$$F(p_1, p_2, \ldots, p_n) = I(A,B) + \lambda\Big(1 - \sum_{i=1}^{n} p_i\Big) = \\ = \sum_{i=1}^{n} p_i \sum_{j=1}^{r} q_{ij} \underbrace{\log_2 \frac{q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}}}_{(*)} + \lambda\Big(1 - \sum_{i=1}^{n} p_i\Big) \,. \quad (5.5)$$

Parciálna derivácia výrazu (*) v (5.5) sa vypočíta takto:

$$\frac{\partial}{\partial p_k} \log_2 \frac{q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}} = \frac{\partial}{\partial p_k} \log_2(e) \cdot \ln \frac{q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}} = \\ = \log_2(e) \cdot \frac{\sum_{t=1}^{n} p_t q_{tj}}{q_{ij}} \cdot \frac{q_{ij}}{-\Big(\sum_{t=1}^{n} p_t q_{tj}\Big)^2} \cdot q_{kj} = -\log_2(e) \cdot \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} \,.$$

Potom pre parciálnu deriváciu funkcie F podľa k-tej premennej platí

$$\begin{aligned} \frac{\partial F}{\partial p_k} &= \frac{\partial}{\partial p_k}\left(I(A,B) + \lambda\Big(1 - \sum_{i=1}^{n} p_i\Big)\right) = \frac{\partial}{\partial p_k}\Big(I(A,B)\Big) - \lambda \\ &= \sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} - \log_2 e \sum_{i=1}^{n} p_i \sum_{j=1}^{r} \frac{q_{ij} q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} - \lambda \\ &= \sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} - \log_2 e \sum_{j=1}^{r} \frac{\sum_{i=1}^{n} p_i q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}} q_{kj} - \lambda \end{aligned} \quad (5.6)$$

$$= \sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} - \log_2 e \sum_{j=1}^{r} q_{kj} - \lambda$$

$$= \sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} - \underbrace{(\log_2 e + \lambda)}_{\gamma} \ . \tag{5.7}$$

Ak označíme $(\log_2 e + \lambda) = \gamma$, položením parciálnych derivácií nule, dostaneme nasledujúcu sústavu rovníc pre neznáme $p_1, p_2, \ldots, p_n$ a γ:

$$\sum_{i=1}^{n} p_i = 1 \tag{5.8}$$

$$\sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} = \gamma \quad \text{pre } k = 1, 2, \ldots, n \ . \tag{5.9}$$

Dá sa ukázať, že funkcia $I(\mathbf{A}, \mathbf{B})$ zo vzťahu (5.1), resp. (5.2) je konkávna, a že splnenie podmienok (5.8) a (5.9) stačí na to, aby príslušná hodnota funkcie $I(\mathbf{A}, \mathbf{B})$ bola maximálna (pozri [7], časť 3.4).
Rovnice (5.8) a (5.9)nazveme **kapacitné rovnice kanála**.
Všimnime si ešte, že ak dosadíme do vzťahu (5.2) na str. 137 pre výpočet $I(\mathbf{A}, \mathbf{B})$ hodnotu γ, ktorá je riešením sústavy (5.8), za

$$\sum_{j=1}^{r} q_{kj} \log_2 \frac{q_{kj}}{\sum_{t=1}^{n} p_t q_{tj}} = \gamma \quad \text{pre } k = 1, 2, \ldots, n \ ,$$

dostaneme

$$I(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} p_i \sum_{j=1}^{r} q_{ij} \log_2 \frac{q_{ij}}{\sum_{t=1}^{n} p_t q_{tj}} = \sum_{i=1}^{n} p_i \gamma = \gamma \sum_{i=1}^{n} p_i = \gamma \ .$$

Ak je γ riešením sústavy (5.8) a (5.9) potom hodnota premennej γ sa rovná maximu informácie, ktoré možno cez kanál preniesť a toto číslo budeme považovať za kapacitu stacionárneho nezávislého kanála. Študujú sa aj všeobecnejšie typy kanálov, rôzne spôsoby definovania ich kapacity sú uvedené v časti 5.5.

Pre symetrický binárny kanál s maticou $\mathbf{Q}$ (5.3) kapacitné rovnice dostanú tvar

$$p_1 + p_2 = 1 \tag{5.10}$$

$$q \log_2 \frac{q}{p_1 q + p_2 (1-q)} + (1-q) \log_2 \frac{1-q}{p_1 (1-q) + p_2 q} = \gamma \tag{5.11}$$

$$(1-q) \log_2 \frac{1-q}{p_1 q + p_2 (1-q)} + q \log_2 \frac{q}{p_1 (1-q) + p_2 q} = \gamma \ . \tag{5.12}$$

Z rovnosti pravých strán (5.11) a (5.12) vyplýva rovnosť ich ľavých strán. Ak od oboch strán tejto rovnosti odčítame $q \log_2 q$ a $(1-q)\log_2(1-q)$, dostaneme

$$\begin{aligned} q\log_2[p_1q + p_2(1-q)] + (1-q)\log_2[p_1(1-q) + p_2q] = \\ = (1-q)\log_2[p_1q + p_2(1-q)] + q\log_2[p_1(1-q) + p_2q] \ , \end{aligned}$$

odkiaľ máme

$$(2q-1)\log_2[p_1q + p_2(1-q)] = (2q-1)\log_2[p_1(1-q) + p_2p] \ . \tag{5.13}$$

Ak by $2q = 1$, v tom prípade $q = 1/2$ a $I(\mathbf{A}, \mathbf{B}) = 0$ bez ohľadu na pravdepodobnosti p_1, p_2.
Ak $q \neq 1/2$ potom z (5.13) máme postupne

$$\begin{aligned} p_1q + p_2(1-q) &= p_1(1-q) + p_2q \\ (2q-1)p_1 &= (2q-1)p_2 \\ p_1 &= p_2. \end{aligned} \tag{5.14}$$

Z (5.10) a (5.14) máme

$$p_1 = p_2 = \frac{1}{2},$$

odkiaľ po dosadení za p_1, p_2 do (5.11) alebo (5.12) máme

$$\gamma = q\log_2(2q) + (1-q)\log_2 2(1-q). \tag{5.15}$$

Symetrický binárny kanál s maticou $\mathbf{Q}$ (5.3) má kapacitu danú vzťahom (5.15) a maximum informácie prenesie pre binárny zdroj s rovnakou pravdepodobnosťou oboch znakov (t. j. pravdepodobnosťou rovnajúcou sa 1/2).

5.4 Množstvo prenesenej informácie

Pripojme na vstup kanála zdroj $\overline{\mathcal{S}} = (Y^*, \mu)$. Pripomeňme, že pravdepodobnosť vyslania slova $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ je $\mu(y_1, y_2, \ldots, y_i)$. Ak sa na vstupe kanála $\mathcal{C} = (Y, Z, \nu)$ budú objavovať vstupné slova zo zdroja $\overline{\mathcal{S}}$, jeho výstup sa bude javiť ako zdroj označovaný symbolom $\mathcal{R} = \mathcal{R}(\mathcal{C}, \overline{\mathcal{S}})$ s abecedou Z a pravdepodobnostnou funkciou π, pre ktorú platí

$$\pi(\mathbf{z}) = \pi(z_1, z_2, \ldots, z_n) =$$
$$= \sum_{\mathbf{y}\in Y^n} \nu(\mathbf{z}|\mathbf{y})\mu(\mathbf{y}) = \sum_{y_1 y_2 \ldots y_n \in Y^n} \nu(z_1, z_2, \ldots, z_n|y_1, y_2, \ldots, y_n)\cdot\mu(y_1, y_2, \ldots, y_n).$$

Okrem výstupného zdroja $\mathcal{R} = \mathcal{R}(\mathcal{C}, \overline{S})$ môžeme dvojici vstupného zdroja $\overline{\mathcal{S}}$ a kanála $\mathcal{C}$ priradiť ešte aj tzv. dvojitý zdroj $\mathcal{D} = ((Y \times Z)^*, \psi))$, ktorý akoby vysielal dvojice (y_i, z_i) vstupného a výstupného znaku. Ak stotožníme slovo $(y_1, z_1)(y_2, z_2)\ldots(y_n, z_n)$ s usporiadanou dvojicou

$$(\mathbf{y}, \mathbf{z}) = ((y_1, y_2, \ldots, y_n), (z_1, z_2, \ldots, z_n)),$$

môžeme pravdepodobnosť

$$\psi\big((y_1, z_1)(y_2, z_2)\ldots(y_n, z_n)\big) = \psi\big((y_1, y_2, \ldots, y_n), (z_1, z_2, \ldots, z_n)\big) = \psi(\mathbf{y}, \mathbf{z})$$

vypočítať nasledovne

$$\psi(\mathbf{y}, \mathbf{z}) = \psi\big((y_1, z_1)(y_2, z_2)\ldots(y_n, z_n)\big) = \psi\big((y_1, y_2, \ldots, y_n), (z_1, z_2, \ldots, z_n)\big) =$$
$$= \nu(\mathbf{z}|\mathbf{y}) \cdot \mu(\mathbf{y}) = \nu(z_1, z_2, \ldots, z_n|y_1, y_2, \ldots, y_n) \cdot \mu(y_1, y_2, \ldots, y_n).$$

Máme teda do činenia s troma zdrojmi – vstupným $\overline{\mathcal{S}}$, výstupným $\mathcal{R} = \mathcal{R}(\mathcal{C}, \overline{S})$ a dvojitým $\overline{\mathcal{D}}$. Fixujme n a označme $\mathbf{A}_n$, $\mathbf{B}_n$ rozklady množiny $Y^n \times Z^n$, na množiny tvaru

$$\{\mathbf{y}\} \times Z^n = \{(y_1, y_2, \ldots, y_n)\} \times Z^n,\ \mathbf{y} = (y_1, y_2, \ldots, y_n) \in Y^n, \quad \text{resp.}$$
$$Y^n \times \{\mathbf{z}\} = Y^n \times \{(z_1, z_2, \ldots, z_n)\},\ \mathbf{z} = (z_1, z_2, \ldots, z_n) \in Z^n,$$

t. j.

$$\mathbf{B}_n = \big\{\{\mathbf{y} \times Z^n\} \mid \mathbf{y} \in Y^n\big\} = \big\{\{(y_1, \ldots, y_n)\} \times Z^n \mid (y_1, \ldots, y_n) \in Y^n\big\}$$
$$\mathbf{A}_n = \big\{\{Y^n \times \mathbf{z}\ \} \mid \mathbf{z} \in Z^n\big\} = \big\{Y^n \times \{(z_1, \ldots, z_n)\} \mid (z_1, \ldots, z_n) \in Z^n\big\}$$

Definujme ešte kombinovaný pokus $\mathbf{D}_n = \mathbf{A}_n \wedge \mathbf{B}_n$. Podľa definície je

$$\mathbf{D}_n = \{(\mathbf{y}, \mathbf{z}) \mid \mathbf{y} \in Y^n, \mathbf{z} \in Z^n\} =$$
$$= \{((y_1, y_2, \ldots, y_n), (z_1, z_2, \ldots, z_n))|(y_1, y_2, \ldots, y_n) \in Y^n, (z_1, z_2, \ldots, z_n) \in Z^n\}.$$

Odpoveď na výsledok pokusu $\mathbf{B}_n$ nám hovorí, aké slovo bolo vyslané. To ale na prijímacej strane kanála $\mathcal{C}$ nevieme. Vieme však výsledok pokusu $\mathbf{A}_n$. Každý

konkrétny výsledok $Y^n \times \{z_1, z_2, \ldots, z_n\}$ pokusu $\mathbf{A}_n$ zmení entropiu $H(\mathbf{B}_n)$ pokusu $\mathbf{B}_n$ na hodnotu $H(\mathbf{B}_n|Y^n \times \{z_1, z_2, \ldots, z_n\})$. Stredná hodnota entropie pokusu $\mathbf{B}_n$ po vykonaní pokusu $\mathbf{A}_n$ je $H(\mathbf{B}_n|\mathbf{A}_n)$. Po vykonaní pokusu $\mathbf{A}_n$ sa teda neurčitosť $H(\mathbf{B}_n)$ zmení na $H(\mathbf{B}_n|\mathbf{A}_n)$.

Rozdiel $H(\mathbf{B}_n) - H(\mathbf{B}_n|\mathbf{A}_n) = I(\mathbf{A}_n, \mathbf{B}_n)$ je stredná informácia o pokuse $\mathbf{B}_n$, ktorú dostaneme po vykonaní pokusu $\mathbf{A}_n$.

Podľa vety 2.13, vzťahu (2.35) (str. 46)je

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$$

Pre náš konkrétny prípad

$$I(\mathbf{A}_n, \mathbf{B}_n) = H(\mathbf{A}_n) + H(\mathbf{B}_n) - H(\mathbf{D}_n)$$

Vieme, že pre entropie vstupného zdroja $\overline{\mathcal{S}}$, výstupného zdroja $\mathcal{R}(\mathcal{C}, \overline{\mathcal{S}})$ a dvojitého zdroja $\overline{\mathcal{D}}$ platí

$$\begin{aligned} H(\overline{\mathcal{S}}) &= \lim_{n\to\infty} \frac{1}{n} \cdot H(\mathbf{B}_n) \\ H(\mathcal{R}) &= \lim_{n\to\infty} \frac{1}{n} \cdot H(\mathbf{A}_n) \\ H(\mathcal{D}) &= \lim_{n\to\infty} \frac{1}{n} \cdot H(\mathbf{D}_n) \end{aligned}$$

Entropia zdroja informácie bola definovaná ako limita stredného množstva informácie pripadajúce na jeden znak pre veľmi dlhé slová. Podobne si môžeme definovať $I(\overline{\mathcal{S}}, \mathcal{R})$ spoločné množstvo informácie vstupného zdroja $\overline{\mathcal{S}}$ a výstupného zdroja $\mathcal{R}$ ako

$$I(\overline{\mathcal{S}}, \mathcal{R}) = \lim_{n\to\infty} \frac{1}{n} \cdot I(\mathbf{A}_n, \mathbf{B}_n) = H(\overline{\mathcal{S}}) + H(\mathcal{R}) - H(\mathcal{D}).$$

Vidíme, že stredné množstvo informácie pripadajúce na jeden znak veľmi dlhých slov prenesených kanálom závisí nielen od vlastností kanála (t. j. podmienených pravdepodobností ν), ale aj od vlastností vstupného zdroja.

5.5 Kapacita kanála

Kapacitu kanála môžeme definovať troma spôsobmi

- pomocou maximálneho množstva informácie pripadajúcej na jeden znak, ktoré je kanál schopný preniesť
- pomocou maximálnej entropie zdroja, ktorého správy je kanál schopný prenášať s ľubovoľne malým rizikom
- pomocou počtu spoľahlivo prenesených postupností.

Tieto tri druhy kapacít si označíme $C_1,\ C_2,\ C_3$.

Kapacita kanála prvého druhu

Kapacitu prvého druhu definujeme nasledovne

$$C_1(\mathcal{C}) = \sup_{\overline{\mathcal{S}}} I(\overline{\mathcal{S}}, \mathcal{R}(\mathcal{C}, \overline{\mathcal{S}})),$$

kde suprémum berieme cez množinu všetkých zdrojov s abecedou Y.

Kapacita kanála druhého druhu

Pre definovanie kapacity druhého druhu potrebujeme najprv definovať celkovú kvalitu prenosu a tiež čo to znamená, že správy zo zdroja $\mathcal{S}$ možno preniesť cez kanál $\mathcal{C}$ s ľubovoľne malým rizikom chyby, krátko rizikom.
V prípade, že vstupná a výstupná abeceda kanála $\mathcal{C}$ sú rovnaké, t. j. ak $Y = Z$ môžeme niekoľkými spôsobmi definovať reálnu funkciu $\mathbf{w}$ s definičným oborom $Y^n \times Z^n$, ktorá pre každú dvojicu slov $\mathbf{y} = y_1 y_2 \dots y_n \in Y^n$, $\mathbf{z} = z_1 z_2 \dots z_n \in Z^n$ vráti reálne číslo $\mathbf{w}(\mathbf{y}, \mathbf{z})$ vyjadrujúce číselne, nakoľko sa prijaté slovo $\mathbf{z}$ líši od vyslaného slova $\mathbf{y}$. Takéto funkcie nazývame váhovými funkciami. Typickými príkladmi váhových funkcií sú dve funkcie $\mathbf{w_e}$ a $\mathbf{w_f}$ definované nasledovne:

$$\mathbf{w_e} = \begin{cases} 0 & \text{ak } \mathbf{y} = \mathbf{z} \\ 1 & \text{inak} \end{cases}$$

$$\mathbf{w_f} = \frac{d(\mathbf{y}, \mathbf{z})}{n}, \quad \text{kde } d \text{ je Hammingova metrika (definícia 4.6, str. 83).}$$

Ak máme kanál $\mathcal{C} = (Y, Z, \nu)$ so zdrojom $\mathcal{S} = (Y^*, \mu)$, potom môžeme pomocou váhovej funkcie $\mathbf{w}$ oceniť kvalitu prenosu ako strednú hodnotu zhody

medzi vyslanou a prijatou postupnosťou

$$\mathbf{r_n}(\mathcal{S}, \mathcal{C}, \mathbf{w}) = \sum_{\mathbf{y} \in Y^n} \sum_{\mathbf{z} \in Z^n} \mathbf{w}(\mathbf{y}, \mathbf{z}) \cdot \nu(\mathbf{z}|\mathbf{y}) \cdot \mu(\mathbf{y}).$$

V prípade úplného prenosového reťazca máme zdroj $\mathcal{S}_X = (X^*, \phi)$, ktorého slová v abecede X zakódujeme zobrazením $h : X^* \to Y^*$ na slová v abecede Y. Vznikne tak zdroj (Y^*, μ), kde $\mu(\mathbf{y}) = 0$, ak neexistuje také slovo $\mathbf{x} \in X^*$, že $\mathbf{y} = h(\mathbf{x})$, inak $\mu(\mathbf{y}) = \phi(h^{-1}(\mathbf{x}))$. Slová zo zdroja (Y^*, μ) sa po prenose kanálom $\mathcal{C}$ objavia na jeho výstupe ako slová v abecedy Z, ktoré nakoniec dekódujeme zobrazením $g : Z^* \to X^*$ na slová v abecede X. Prenos slova $\mathbf{x} \in X^n$ bude teda nasledovný

$$\begin{aligned} \mathbf{x} \in X^n \to \mathbf{y} = h(\mathbf{x}) \in Y^n \to & \text{ vstup do kanála } \mathcal{C} \to \\ & \to \text{ výstup z kanála } \mathcal{C} \to \mathbf{z} \in Z^n \to g(\mathbf{z}) \in X^n \end{aligned}$$

Po vyslaní slova $\mathbf{x} \in X^n$ prijmeme slovo $g(\mathbf{z})$, pričom prípadnú odchýlku ohodnotíme ako $\mathbf{w}(\mathbf{x}, g(\mathbf{z}))$. Celkovú kvalitu prenosu ohodnotíme nasledovne:

$$\begin{aligned} \mathbf{r_n}(\mathcal{S}_X, h, \mathcal{C}, g, \mathbf{w}) &= \sum_{\mathbf{x} \in X^n} \sum_{\mathbf{z} \in Z^n} \mathbf{w}(\mathbf{x}, g(\mathbf{z})) \cdot \nu(\mathbf{z}|h(\mathbf{x})) \cdot \mu(h(\mathbf{x})) \\ &= \sum_{\mathbf{x} \in X^n} \sum_{\mathbf{z} \in Z^n} \mathbf{w}(\mathbf{x}, g(\mathbf{z})) \cdot \nu(\mathbf{z}|h(\mathbf{x})) \cdot \phi(\mathbf{x}) \ . \end{aligned}$$

Hodnotu $\mathbf{r}_n$ nazývame **rizikom**. Ak je riziko malé, prenos je bez veľkého počtu chýb. Naopak, ak je riziko $\mathbf{r}_n$ veľké, pri prenose sa objavuje veľký počet nesprávne prijatých slov dĺžky n.

Definícia 5.2. Hovoríme, že pri danej váhovej funkcii $\mathbf{w}$ môžeme správy zo zdroja $\mathcal{S}_X = (X, \phi)$ preniesť cez kanál $\mathcal{C} = (Y, z, \nu)$ **s ľubovoľne malým rizikom**, ak k ľubovoľnému ε existuje také n a také kódové a dekódové zobrazenia h a g, že $\mathbf{r_n}(\mathcal{S}_X, h, \mathcal{C}, g, \mathbf{w}) < \varepsilon$.

Definícia 5.3. Definujme

$$C_2^e = \sup_{\mathcal{S}} H(\mathcal{S}), \qquad C_2^f = \sup_{\mathcal{S}} H(\mathcal{S}),$$

kde suprémum v oboch prípadoch berieme cez množinu všetkých zdrojov $\mathcal{S}$ prenesiteľných cez kanál $\mathcal{C}$ s ľubovoľne malým rizikom a kde kladieme $\mathbf{w} = \mathbf{w}_e$ pre C_2^e a $\mathbf{w} = \mathbf{w}_f$ pre C_2^f.

Kapacita kanála tretieho druhu

Pri definícii kapacity kanála tretím spôsobom vychádzame z nasledujúceho pojmu ε-rozlíšiteľnosti slov.

Definícia 5.4. Množina $U \subseteq Y^n$ vstupných slov je **ε-rozlíšiteľná**, ak existuje taký rozklad $\{Z(\mathbf{u}) : \mathbf{u} \in U\}$ množiny Z^n, že

$$\nu(Z(\mathbf{u})|\mathbf{u}) \geq 1 - \varepsilon.$$

Pripomeňme, že rozklad $\{Z(\mathbf{u}) : \mathbf{u} \in U\}$ je systém takých podmnožín množiny Z^n, že platí:

1. Pre $\mathbf{u},\ \mathbf{v} \in U$, $\mathbf{u} \neq \mathbf{w}$ je $Z(\mathbf{u}) \cap Z(\mathbf{v}) = \varnothing$
2. $\bigcup_{\mathbf{u} \in U} Z(\mathbf{u}) = Z^n$.

Číslo $\nu(Z(\mathbf{u})|\mathbf{u})$ je podmienená pravdepodobnosť javu, že ak vyšleme slovo $\mathbf{u}$, k nemu prijaté slovo padne do množiny $Z(\mathbf{u})$.

Ak množina $U \subseteq Y^n$ vstupných slov je ε-rozlíšiteľná, a my prijmeme slovo z množiny $Z(\mathbf{u})$, vieme, že s pravdepodobnosťou $1 - \varepsilon$ bolo vyslané slovo $\mathbf{u}$.

Pre $\varepsilon > 0$, kanál $\mathcal{C}$ a prirodzené číslo n označme $d_n(\mathcal{C}, \varepsilon)$ maximálny počet ε-rozlíšiteľných slov z Y^n. Potom tretí druh C_3 kapacity kanála $\mathcal{C}$ definujeme takto

$$C_3(\mathcal{C}) = \inf_{\varepsilon} \limsup_{n \to \infty} \frac{1}{n} \log_2 d_n(\mathcal{C}, \varepsilon).$$

Dá sa ukázať, že pre väčšinu kanálov $\mathcal{C}$, ktoré majú praktický význam platí

$$C_1(\mathcal{C}) = C_2^e(\mathcal{C}) = C_2^f(\mathcal{C}) = C_3(\mathcal{C}),$$

z čoho vyplýva, že všetky definície kapacity kanálu boli zvolené zmysluplne.

Prístup k trom definíciám kapacity kanálov bol prevzatý z práce [5]. Iný prístup s analogickými výsledkami nájde čitateľ v knižke [9].

5.6 Shannonove vety

V tejto časti budeme uvažovať zdroj $\mathcal{S}$ s entropiou $H(\mathcal{S})$ a prenosový kanál $\mathcal{C}$ s kapacitou $C(\mathcal{C})$.

Veta 5.1 (Priama Shannonova veta)**.** *Ak pre stacionárny nezávislý zdroj $\mathcal{S}$ a pre stacionárny nezávislý kanál $\mathcal{C}$ platí*

$$H(\mathcal{S}) < C(\mathcal{C}),$$

potom možno správy zdroja $\mathcal{S}$ preniesť cez kanál $\mathcal{C}$ s ľubovoľne malým rizikom.

Veta 5.2 (Obrátená Shannonova veta)**.** *Ak pre stacionárny nezávislý zdroj $\mathcal{S}$ a pre stacionárny nezávislý kanál $\mathcal{C}$ platí*

$$H(\mathcal{S}) > C(\mathcal{C}),$$

potom nemožno správy zdroja $\mathcal{S}$ preniesť cez kanál $\mathcal{C}$ s ľubovoľne malým rizikom.

Shannonove vety platia aj pre oveľa všeobecnejšie triedy zdrojov a kanálov napr. pre ergodické zdroje a ergodické kanály. Shannonove vetu ukazujú, že pojmy entropie zdroja a kapacity kanála boli zvolené dobre a navzájom úzko súvisia. Dôkazy Shannonových viet nájde čitateľ napríklad v [9], niektoré jej varianty v [3]. Vzhľadom k rozsahu a určeniu tejto publikácie ich tu neuvádzam.

Register

Literatúra

[1] Adámek, J.: *Kódování*, SNTL Praha, 1989

[2] Berlehamp, R., R.: *Algebraic Coding Theory*, McGraw-Hill, New York, 1968 (*Ruský preklad: Algebrajičeskaja teorija kodirovanija, Mir, Moskva, 1971*)

[3] Billingsley, P.: *Ergodic Theory and Information*, J. Willey and Sons, Inc., New York, London, Sydney, 1965 (*Ruský preklad: Ergodičeskaja teorija i informacija, Mir, Moskva, 1969*)

[4] Černý, J., Brunovský, P.: *A Note on Information Without Probablity*, Information and Control, pp. 134 - 144, Vol. 25, No. 2, June, 1974

[5] Černý, J.: *Entropia a informácia v kybernetike*, Alfa – vydavateľstvo technickej a ekonomickej literatúry, Bratislava, 1981

[6] Halmos, P., R.: *Measure Theory (Graduate Texts in Mathematics)*, Springer Verlag,

[7] Hankerson, D., Harris, G.,A., Johnson, O.,D., Jr.: *Introduction to Information Theory nad Data Compression*, CRC Press LLC, 1998, ISBN 0-8493-3985-5

[8] Jaglom, A., M., Jaglom, I., M.: *Pravděpodobnost a informace*, ČSAV, Praha, 1964

[9] Kolesnik, V., D., Poltyrev, G., Š.,: *Kurs teorii informacii*, Nauka, Moskva, 1982

[10] Neubrunn, T., Riečan, B.,: *Miera a integrál, Veda, Bratislava, 1981*

[11] Schulz, R., H.: *Codierungstheorie, Eine Einfuhrung*, Vieweg, Wiesbaden 1991, ISBN 3-528-06419-6

Autor:	Doc. RNDr. Stanislav Palúch, CSc.
Názov:	**Teória informácie**
Vydala:	Žilinská univerzita v EDIS-vydavateľstve ŽU v xxxxx 2008 ako svoju XXX. publikáciu
Zodpovedný redaktor:	XXXXXX
Technický redaktor:	XXXXXX
Vydanie:	prvé
Náklad:	XXX výtlačkov
AH/VH:	XX/XX
Druh tlače:	ofset
Typografický systém:	LaTeX pod o. s. Linux
ISBN 80-XXXX-XX-X	