Úvod

So vstupom do tretieho tisícročia vstupuje ľudstvo do veku, ktorý sa bude právom charakterizovať prívlastkom *informačný*. Sme a stále viac budeme zahrňovaní množstvom informáciií 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 neautorizovnou 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 ifnormácie medzi široké vrstvy ľudí spôsobilo enormný nárast vzdelanosti a dalo základy pre priemyslenú 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 ukladacích 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.

Uvádza sa, že pád Rímskej ríše mal viac príčin – od vnútorného rozkladu spoločnosti až po nájazdy barbarov, ale aj to, že vládny informačný systém nebol adekvátny rozľahlosti územia a hierarchickej štruktúre politickej organizácie ríše.

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 discilíny. Musel teda vzniknúť nový vedný odbor – teória informácie. Tá sa stala súčasťou vtedy vznikajúcej vedy o riadení – matematickej kybernetiky, ktorá postupným vývojom vrástla 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

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 prirodzene definovať ako maximálne množstvo informácie, ktoré môže preniesť za jednotku času.

Na to však treba kvantifikovať informáciu, t.j. určiť, ako ju merať a definovať jej jednotkové množstvo. 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 ztotožniť s dátami, ktoré predstavujú jej zápis.

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 fm.vse.cz). 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 treba charakterizovať a kvantitatívne ohodnotiť informačnú výdatnosť zdroja – jeho *entropiu*.

Potom už mala teória informácie jasne vytýčenú cestu k jednému zo svojich najdôležitejších výsledkov – Shannonovej vete: *Ak je entropia zdroja menšia než kapacita kanála, potom vždy existuje kódovanie, ktoré umožní preniesť správy zo zdroja cez kanál bez zdržania s dostatočnou presnosťou.* Obrátená veta zase tvrdí, že ak je entropia zdroja väčšia než kapacita kanála, takýto prenos nie je možný.

Kapitola 1

Informácia

1.1 Možnosti a spôsoby zavedenia informácie

Ak žiadame informáciu o odchode rýchlika Tatran zo Žiliny do Bratislavy, môžeme ju dostať v presnej forme nasledujúcej vety: „Rýchlik Tatran do Bratislavy odchádza zo Žilliny o 20 hodine 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 dvojku.“ 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 2“ 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 existujú v našom hodnotiacom systéme len 4 možnosti. 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 predpokladť, že informácia bude definovaná ako reálna funkcia, 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 štandartnejšími matematickými objektami. Každý výrok nesúci informáciu je vlastne veta v tvare: „Nastal jav A.“ resp. „Nastane jav A.“ Tu možno jav A považovať za podmnožinu istého základného priestoru Ω, presnejšie za prvok σ-algebry $\mathcal{A}$ podmnožín základného priestoru Ω.

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

$$1.\quad \Omega \in \mathcal{A} \tag{1.1}$$

$$2.\quad \text{Ak } A \in \mathcal{A} \text{ potom aj } A^C = (\Omega - A) \in \mathcal{A} \tag{1.2}$$

$$3.\quad \text{Ak } A_n \in \mathcal{A} \text{ pre } n = 1, 2, \ldots, \infty, \text{ potom aj } \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}. \tag{1.3}$$

σ-algebra teda obsahuje základný priestor Ω, s každou 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 postupnosťou množín obshuje aj ich spoločný prienik.

Ak teda vezmeme za množinu $\mathcal{A}$ σ-algebru podmnožín nejkého základného priestoru, máme prvý problém vyriešený. Druhým problémom je, ako zaviesť reálnu funkciu $I : \mathcal{A} \to \mathbb{R}$ (kde $\mathbb{R}$ je množinareálnych čísel) tak, aby hodnota $I(A)$ pre $A \in \mathcal{A}$ vyjadrovala informáciu, ktorú dostaneme v správe, že nastal jav A.

V analogickej sitácii sme boli, keď sme zavádzali pravdepodobnosť na σ-algebre $\mathcal{A}$. Tu sa dalo postupovať trojako:
a) elementárne. Predpokladáme, že Základný priestor Ω pozostáva z konečného počtu n rovnako pravdepodobných elementárnych javov; pravdepodobnosť každého z nich musí byť rovná $\frac{1}{n}$. Za prvky σ-algebry $\mathcal{A}$ berieme $\emptyset$ a všetky konečné zjednotenia elementárnych javov. Potom každej množine $A \in \mathcal{A}$ priradíme pravdepodobnosť $\frac{m}{n}$, kde m je počet elementárnych javov obsažených v tejto množine. Tento postup možno zovšeobecniť aj pre prípad, kedy elementárne javy majú nerovnaké pravdepodobnosti $p_1, p_2, \ldots, p_n$, kde $p_1 + p_2 + \cdots + p_n = 1$.
b) axiomaticky
c) pomocou známych pojmov – normovanej miery na merateľnom priestore $(\Omega, \mathcal{A})$

Základnou vlastnoťou 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 ifnormáciu $I(A)$ 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 $\oplus : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, pomocu ktorej bude platiť $I(A \cup B) = I(A) \oplus I(B)$ pre ľubovoľné $A \in \mathcal{A}$, $B \in \mathcal{A}$, $A \cap B = \emptyset$. Pochopiteľne, zatiaľ nevieme, či vôbec taká operácia existuje, a ak existuuje, či ich je viac a ako sa od seba líšia.

Spíšme si, aké vlastnosti očakávame od informácie

1. $I(A) \geq 0$ pre všetky $A \in \mathcal{A}$ (1.4)
2. $I(\Omega) = 0$ (1.5)
3. Ak $A \in \mathcal{A},\ B \in \mathcal{A},\ A \cap B = \emptyset,$ potom $I(A \cup B) = I(A) \oplus I(B)$ (1.6)
4. Ak $A_n \to A,$ potom $I(A_n) \to A.$ (1.7)

Vlastnosť 1. hovorí, že množstvo informácie je nezáporné číslo, vlastnosť 2. hovorí, že z toho, že nastal jav Ω nezískame žiadnu informáciu. Vlastnosť 3. hovrí, že informáciu disjunktného zjednotenia javov dostaneme z informácií jednotlivých javovo pomocou operácie $\oplus$ a posledná 4. vlastnosť hovorí, že ak $A = \bigcup_{n=1}^{\infty} A_n$ alebo $A = \bigcap_{n-1}^{\infty} A_n$, potom $I(A) = \lim_{n\to\infty} I(A_n)$, t.j. 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 $I(A \cap B)$ rovná súčtu informácií oboch javov. Z toho nasledujúca definícia:

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

$$I(A \cap B) = I(A) + I(B) \tag{1.8}$$

Teraz zosumarizujeme vlastnosti binárnej operácie $\oplus$

1. $x \oplus y = y \oplus x$ (1.9)
2. $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ (1.10)
3. $I(A) \oplus I(A^C) = 0$ (1.11)
4. $\oplus : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ je spojitá funkcia dvoch premenných (1.12)
5. $(x + z) \oplus (y + z) = (x \oplus y) + z$ (1.13)

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 sa z toho, že nastal jav A nič nedozviem o jave C a ani z toho, že nastal jav B 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 $I((A \cup B) \cap C))$

$$I((A \cup B) \cap C)) = I(A \cup B) + I(C) = I(A) \oplus I(B) + I(C) = x \oplus y + z \tag{1.14}$$

$$\begin{aligned}((A \cup B) \cap C)) &= I((A \cap C) \cup (B) \cap C)) = \\ &= I(A \cap C) \oplus I(B \cap C) = ((I(A) + I(C)) \oplus ((I(B) + I(C)) = (x + z) \oplus (y + z)\end{aligned} \tag{1.15}$$

Porovnaním vzťahov (1.14), (1.15) dostaneme žiadanú vlastnosť 5.

Veta 1.1. *Nech binárna operácia $\oplus$ vyhovuje axiómam* (1.9) *až* (1.13). *Potom*

buď $x \oplus y = \min\{x, y\}$, (1.16)

alebo $x \oplus y = -k \log_2 \left(2^{-\frac{x}{k}} + 2^{-\frac{y}{k}}\right)$, *kde* $k > 0$. (1.17)

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

Zaujímavé ale je, že (1.16) je limitným prípadom (1.17) 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

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

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

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

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

$$1. \quad \Omega = \bigcup_{i=}^{n} A_i, \text{ kde } A_i \cap A_j = \emptyset \text{ pre } i \neq j \tag{1.18}$$

$$2. \quad I(A_1) = I(A_2) = \cdots = I(A_n) = a \text{ pre } i \neq j \tag{1.19}$$

Chceme určiť veličinu a. Z (1.18), (1.19) vyplýva

$$0 = I(\Omega) = I(A_1) \oplus I(A_2) \oplus \cdots \oplus I(A_n) = \underbrace{a \oplus a \oplus \cdots \oplus a}_{n-\text{krát}} = \bigoplus_{i=1}^{n} a \tag{1.20}$$

$$0 = \bigoplus_{i=1}^{n} a = \begin{cases} \min\{a, a, \ldots, a\} = a & \text{ak } x \oplus y = \min\{x, y\} \\ -k.\log_2\left(2^{-a/k} + 2^{-a/k} + \cdots + 2^{-a/k}\right) & \text{ak } x \oplus y = -k\log_2\left(2^{-x/k} + 2^{-y/k}\right) \end{cases} \tag{1.21}$$

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

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

Z posledného vyťahu vyplýva, že

$$a = k.\log_2(n) = -k.\log_2\left(\frac{1}{n}\right) \tag{1.22}$$

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

$$I(A) = I(A_{i_1}) \oplus I(A_{i_2}) \oplus \cdots \oplus I(A_{i_m}) = \underbrace{a \oplus a \oplus \cdots \oplus a}_{m-\text{krát}} = -k.\log_2\left(\underbrace{2^{-a/k} + 2^{-a/k} + \cdots + 2^{-a/k}}_{m-\text{krát}}\right) =$$
$$= -k\log_2\left(m.2^{-a/k}\right) = -k.\log_2(m) - k.\log_2\left(2^{-a/k}\right) = -k.\log_2(m) - k.(-a/k) =$$
$$= -k.\log_2(m) + a = -k.\log_2(m) + k.\log_2(n) = k.\log_2\left(\frac{n}{m}\right) = -k.\log_2\left(\frac{m}{n}\right) \tag{1.23}$$

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 (1.21), 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 $\log_b(x) = \log_b(a).\log_a(x) = (1/\log_a(b)).\log_a(x)$ Namiesto konštanty k a logaritmu pri základe 2 by mohol vo vzťahoch (1.22), (1.23) vystupovať iba logaritmus pri ľubovoľnom základe. Toto skutočne niektorí autori aj používajú, najmä v staršej literatúre sa občas objaví používanie dekadického logaritmu.

Pre určenie koštanty k môže byť užitočná nasledujúca úvaha. Výpočtová a digitálne 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 (1.22) $1 = a = k.\log_2(2) = k$. Ak teda chceme, aby vzorce (1.22), (1.23) vyjadrovali množstvo ifnormácie v bitoch, musíme v nich položiť $k = 1$. Odteraz budeme predpokladať, že informáciu meriame v bitoch a teda že k=1.

1.3 Axiomatická definícia informácie

Definícia 1.3. Merateľný priestor $\mathcal{P} = (\Omega, \mathcal{A})$ je usporiadaná dvojica, kde Ω je množina nazývaná tiež základný priestor, a $\mathcal{A}$ je σ-algebra podmnožín základného priestoru Ω.

S využitím operácie $\oplus$ môžeme axiomaticky definovať informáciu I na merateľnom priestore $(\Omega, \mathcal{A})$ celkom analogicky ako pravdepodobnosť.

1. I je definovaná na merateľnom priestore$(\Omega, \mathcal{A})$ (1.24)
2. $I(\emptyset) = \infty, \quad I(\Omega) = 0$ (1.25)
3. $I\left(\bigcup_n A_n\right) = I(A_1) \oplus I(A_2) \oplus \cdots = \bigoplus_n I(A_n)$ pre postupnosť disjunktných javov $\{A_n\}$ (1.26)

Podobne ako v teórii pravdepodobnosti sa dá definovať nezávislosť javov.

Definícia 1.4. Konečná alebo nekonečná postupnosť $\{A_n\}_n$ sa nazýva postupnosťou nezávislých javov, ak pre každú konečnú vybranú postupnosť $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ platí

$$I\left(\bigcap_{k=1}^{m} A_{i_k}\right) = \sum_{k=1}^{m} I\left(A_{i_k}\right) \tag{1.27}$$

1.4 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 pripadjúcou do úvahy je funkcia $f(x) = -k.\log_2(x)$. Použijeme pritom postu podľa [Černého].

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 pre postupnosť nezávislých javov $A_1, A_2, \ldots, A_n$ platí

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

a súčasne

$$I(A_1 \cap A_2 \cap \cdots \cap A_n) = \sum_{i=1}^{n} I(A_i) = \sum_{i-1}^{n} f\left(P(A_i)\right) \tag{1.29}$$

Ľavé strany posledných dvoch vzťahov musia byť rovné, a preto

$$f\left(\prod_{i-1}^{n} P(A_i)\right) = \sum_{i-1}^{n} f\left(P(A_i)\right) \tag{1.30}$$

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

$$f\left(\frac{1}{2^x}\right) = x.f\left(\frac{1}{2}\right) \text{ pre všetky reálne čísla } x \in \langle 0, \infty) \tag{1.33}$$

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

$$\begin{aligned} g(x) = f(x) + f\left(\frac{1}{2}\right).\log_2(x) &= f\left(2^{\log_2(x)}\right) + f\left(\frac{1}{2}\right).\log_2(x) = \\ &= f\left(\frac{1}{2^{-\log_2(x)}}\right) + f\left(\frac{1}{2}\right).\log_2(x) = -\log_2(x).f\left(\frac{1}{2}\right) + f\left(\frac{1}{2}\right).\log_2(x) = 0 \end{aligned} \tag{1.34}$$

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

$$f(x) = -f\left(\frac{1}{2}\right).\log_2(x) = -k.\log_2(x) \tag{1.35}$$

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

$$I(A) = -k.\log_2(P(A)) \tag{1.36}$$

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 inak sa to proste nedá.

Kapitola 2

Entropia

2.1 Shannonova definícia entropie

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 pokus je organizovaný tak, že sa po jeho vykonaní dozvieme, ktorý z týchto disjunktných javov nastal. Teraz sa môžeme sa spýtať, akú neistotu máme pred vykonaním tohoto pokusu, alebo akú informáciu získame po jeho vykonaní.

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.

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 je výsledok pokusu $\mathbf{P}_1$ dostaneme informáciu 10.49 bitov, podobne po vykonaní pokusu $\mathbf{P}_2$ dostanem vždy informáciu 1 bit. Hovoríme, že entropia $H(\mathbf{P}_1)$ pokusu $\mathbf{P}_1$ je 10.49 bitov, entropia $H(\mathbf{P}_2)$ pokusu $\mathbf{P}_2$ je 1 bit. Vieme teda definovať neurčitosť - entropiu 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 pokusie $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$ 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 dodefinovať funkciu $\eta(x) = -x.\log_2(x)$ pre $x = 0$ ako $\eta(0) = 0$. Potom by Shannonova formula pre entropiu mala byť v tvare $H(\mathbf{P}) = \sum_{i=1}^{n} \eta(P(A_i))$, kde $\eta(x) = -x.\log_2(x)$. 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 prepokladať 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 $\mathbf{P}$.

2.2 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 nieko sústav axióm, my uvedieme tzv. Fadejevovu sústavu z roku 1956:

AF0: $H(p_1, p_2, \ldots, p_n)$ 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: $H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1\rangle$.

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

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

[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 $\chi_{A_i}(\omega) = 0$.

AF3: Ak $p_n = q_1 + q_2 > 0$, $q_{\geq}0$, $q_2 \geq 0$, 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 jedho 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 poslednú množinu 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

$$H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right). \tag{2.5}$$

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

$$p_n.H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right). \tag{2.6}$$

Fadejevove axiómy AF0, až AF3 sú dostatočné na odvodenie tvaru funkcie H a dá sa z nich dokázať i platnosť Shannonovho axiómu 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 axiómov 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 formuli. Pretože podľa vety 2.1 Shannonova entropia daná vzorcom (2.1) spĺňa všetky axiómy, nasledujúce vety platia aj pre ňu.

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

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

Dôkaz. Matematickou indukciou podľa m. Pre $m = 2$ je tvrdenie axiómou AF1. Nech funkcia $H(x_1, x_2, \ldots x_m)$ 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:

$$H(p_1, p_2, \ldots, p_m, p_{m+1}) =$$
$$= H(p_1, p_2, \ldots, p_{m-1}, (p_m + p_{m+1})) + (p_m + p_{m+1}).H\left(\frac{p_m}{(p_m + p_{m+1})}, \frac{p_{m+1}}{(p_m + p_{m+1})}\right) \quad (2.7)$$

Spojitosť prvého sčítanca pravej strany (2.7) vyplýva z indukčného predpokladu, spojitosť druhého sčítanca vyplýva z axiómy AF1. □

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

Dôkaz.

$$H\left(\frac{1}{2}, \underbrace{\frac{1}{2}, 0}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}.H(1, 0) \quad (2.8)$$

$$H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(0, \underbrace{\frac{1}{2}, \frac{1}{2}}\right) = H(0, 1) + H\left(\frac{1}{2}, \frac{1}{2}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + H(1, 0) \quad (2.9)$$

Porovnaním pravých strán (2.8), (2.9) 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 čísliel $p_1, p_2, \ldots, p_n$ je kladné. Nech je to p_n (inak zmeníme poradie). Potom

$$H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n) + p_n.\underbrace{H(1, 0)}_{0} \quad (2.10)$$

□

Zase jedna dobrá vlastnoť 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*

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) = H(p_1, p_2, \ldots, p_n) + p_n.H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right) \quad (2.11)$$

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

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, \underbrace{q_2, \ldots, q_{m+1}}_{p'}) = H(p_1, p_2, \ldots, p_{n-1}, \underbrace{q_1, p'}_{p_n}) + p'.H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right) =$$
$$H(p_1, p_2, \ldots, p_n) + p_n.\left[H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n}H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right)\right] \quad (2.12)$$

Ďalej podľa indukčného predpokladu platí

$$H\left(\frac{q_1}{p_n}, \underbrace{\frac{q_2}{p_n}, \ldots, \frac{q_{m+1}}{p_n}}_{\frac{p'}{p_n}}\right) = H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n} H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right) \tag{2.13}$$

Vidíme, že pravá strana (2.13) je totožná s obsahom veľkej hranatej zátvorky na pravej strane vzťahu (2.12). Dosadením ľavej strany vyťahu (2.13) do (2.12) dostávame (2.11). □

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

$$H(q_{11}, q_{12} \ldots q_{1m_1}, q_{21}, q_{22}, \ldots, q_{2m_2}, \ldots, q_{n1}, q_{n2}, \ldots, q_{nm_n}) = \\ = H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i . H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \ldots, \frac{q_{im_i}}{p_i}\right) \tag{2.14}$$

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.

$$F(mn) = H\left(\underbrace{\underbrace{\frac{1}{mn}, \ldots, \frac{1}{mn}}_{m\text{-krát}}, \ldots \underbrace{\frac{1}{mn}, \ldots, \frac{1}{mn}}_{m\text{-krát}}}_{n\text{-krát}}\right) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + \sum_{i-1}^{n} \frac{1}{n} H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = \\ = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = F(n) + F(m) \tag{2.15}$$

□

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. 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:

$$F(n^k) = k.F(n) \tag{2.16}$$

Vzťah (2.16) 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 a ľubovoľne veľké prirodzené číslo K. Potom existuje prirodzené číslo k také, že

$$m^k \leq n^K < m^{k+1} \tag{2.17}$$

Pretože F je rastúca funkcia, je aj

$$F(m^k) \leq F(n^K) < F(m^{k+1}) \quad \text{a pretože (2.16)} \quad k.F(m) \leq K.F(n) < (k+1).F(m)$$

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

$$\frac{k}{K} \leq \frac{F(n)}{F(m)} < \frac{k+1}{K} \tag{2.18}$$

Pretože je (2.17) môžeme podobnou úvahou písať

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

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

$$\frac{k}{K} \leq \frac{\log_2(n)}{\log_2(m)} < \frac{k+1}{K} \tag{2.19}$$

Ak porovnáme výrazy (2.18) a (2.19) 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

$$\left| \frac{F(n)}{F(m)} - \frac{\log_2(n)}{\log_2(m)} \right| < \frac{1}{K} \tag{2.20}$$

Celý postup môžeme zopakovať pre ľubovoľne veľké číslo K a preto (2.20) musí platiť pre ľubovľné K, čo je možné len tak, že

$$\frac{F(n)}{F(m)} = \frac{\log_2(n)}{\log_2(m)} \quad \text{a teda} \quad F(n) = F(m).\frac{\log_2(n)}{\log_2(m)} = \left(\frac{F(m)}{\log_2(m)} \right) \log_2(n) \tag{2.21}$$

Ak v (2.21) 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*

$$H(p_1, p_2, \ldots, p_n) = -c.\sum_{i=1}^{n} p_i.\log_2(p_i) \tag{2.22}$$

Dôkaz. Dokážeme najprv (2.22) 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 (2.14) vety 2.6 môžeme písať

$$H\left(\underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_1\text{-krát}}, \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_2\text{-krát}}, \ \ldots\ \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{q_n\text{-krát}} \right) = H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i.H\left(\frac{1}{q_i}, \frac{1}{q_i} \ldots, \frac{1}{q_i} \right) =$$

$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i.F(q_i) = H(p_1, p_2, \ldots, p_n) + c.\sum_{i=1}^{n} p_i.\log_2(q_i) \tag{2.23}$$

Pretože ľavá strana (2.23) je rovná $F(s) = c.\log_2(s)$, môžeme písať

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

$$= -c\sum_{i=1}^{n} p_i.(\log_2(q_i) - \log_2(s)) = -c\sum_{i=1}^{n} p_i.\log_2\left(\frac{q_i}{s} \right) = -c.\sum_{i=1}^{n} p_i.\log_2(p_i) \tag{2.24}$$

Pretože funkcia H je spojitá a (2.24) platí pre vštky racionálne $p_1 \geq 0,\ p_2 \geq 0, \ldots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, musí (2.24) platiť aj pre všetky reálne argumenty p_i splňujúce 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 formuli, ako v prípade, keď sme entropiu definovali ako strednú hodnotu diskrétnej náhodnej veličiny informácie.

2.3 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*

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} q_i \log_2(p_i) \tag{2.25}$$

Dôkaz. Najprv dokážeme platnosť nerovnosti

$$\ln(1+y) \leq y \quad \text{pre} \quad y > -1 \tag{2.26}$$

Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy. Je $g'(y) = \dfrac{1}{y} - 1$, $g''(y) = -\dfrac{1}{y^2} \geq 0$, rovnica $g'(y) = 0$ má jediné riešenie $y = 1$ a $g'(1) = -1 < 0$. Funkcia $g(y)$ nadobúda svoje maximum v bode $y = 1$. Je preto $g(y) \leq 1$, t.j. $\ln(1+y) - y \leq 1$ a teda $\ln(1+y) \leq y$, pričom rovnosť nastáva práve vtedy, keď $y = 1$. Ak v (2.25) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\ln(x) \leq x - 1 \quad \text{pre} \quad x > 0, \tag{2.27}$$

pričom rovnosť nastáva práve vtedy, keď $x = 0$. Dosaďme teraz do (2.27) za $x = \dfrac{q_i}{p_i}$

$$\begin{aligned}
\ln(q_i) - \ln(p_i) &\leq \frac{q_i}{p_i} - 1 \\
p_i \ln(q_i) - p_i \ln(p_i) &\leq q_i - p_i \\
-p_i \ln(p_i) &\leq -p_i \ln(q_i) + q_i - p_i \\
-\sum_{i=1}^{n} p_i \ln(p_i) &\leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1} \\
-\sum_{i=1}^{n} p_i \frac{\ln(p_i)}{\ln(2)} &\leq -\sum_{i=1}^{n} p_i \frac{\ln(q_i)}{\ln(2)} \\
-\sum_{i=1}^{n} p_i \log_2(p_i) &\leq -\sum_{i=1}^{n} p_i \log_2(q_i),
\end{aligned}$$

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 a položme v (2.25) $q_1 = q_2 = \cdots = q_n = \frac{1}{n}$. Potom

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

□

2.4 Použitie entropie pri riešení niektorých úloh

Veta 2.11 spolu s axiómou AS4 má tento dôsledok pre organizovanie pokusu: Ak chceme pokusom získať čo najväčšiu informáciu, snažme sa ho usporiadať tak, aby mal čo najväčší počet rovnako pravdepodobných výsledkov. Často sa vyskytujú úlohy typu: „Zistite na najmenší počet meraní (otázok, skúšok), ktorý elementárny jav konečeného pravdepodobnostného priestoru ω nastal." Najlepšie by bolo, keby sa dal pokus naplánovať tak, aby dal toľko možných odpovedí, koľko je možných výsledkov.

Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. Predpokladajme, že nastal jav $\omega \in \Omega$. My nemáme možnosť (a ani potrebu) zistiť, o ktorý jav ide, nám stačí vedieť do ktorej množiny disjunktného rozkladu $\mathbf{B} = \{B_1, B_2, \ldots, B_n\}$ základného 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. Veľmi často je $(\Omega, \mathcal{A}, P)$ konečný pravdepodobnostný priestor $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. a základný pokus má tvar $\mathbf{B} = \{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}\}$.

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 nemôžeme zorganizovať priamo základný pokus $\mathbf{B}$, 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 - „ano" 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. 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 $H(1/32, 31/32) = -(1/32).\log_2(1/32) - (31/32).\log_2(31/32) = -(1/32).(-5) - (31/32).(-0.0458) = 0.15625 + 0.04437 = 0.20062$ 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

[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.

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 pripadajúch do úvahy, na dve rovnaké polovice. Potom je možné dopracovať sa výsledku po piatich otázkach.

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?“. Výnimkou je pacient na stomatochirurgii so zadrôtovanými ústami po fraktúre sánky schopný len prikývnuť alebo zvrtieť hlavou.

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.

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, 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 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. $|A| = |B| = m + 1$, $|C| = m$.

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, 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.

Majme n veľkých bední s guličkami. Guličky vo všetkých bedniach sú rovnaké až na jednu, v ktorej sú guličky o 1 gram ťažšie, než v ostatných bedniach. Máme k dispozícii presné dvojmiskové váhy so závažím, ktorými dokážeme odvážiť ľubovoľné množstvo guličiek s presnosťou lepšou než 1

gram. Na koľko vážení možno zaručene určiť, v ktorej bedni sú ťažšie guličky? Základný pokus má n možných výsledkov, jeho entropia je $\log_2(n)$. Ak by sme z každej bedne vybrali po guličke, 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 prvej bedne dáme na ľavú misku 1 guličku, z druhej 2, atď. až z n-tej bedne n. Na pravú misku dáme $1+2+\cdots+n = (1/2)n(n+1)$ guličiek z prvej bedne. Ak preváži pravá miska, ťažšie guličky sú v prvej debne. Ak preváži ľavá miska, dodáme závažie k gramov na vyváženie misiek. Potom ťažšie guličky sú v k-tej debne. Na vyriešenie úlohy stačí jedno váženie.

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, $Y \in \langle 0, 1000\rangle$. 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 vdenie 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

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

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 apoň 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.

Postup pri určovaní jedného z n možných elementárnych javov sme robili pomocou niekoľkých pokusov nasledovne: V prvom kroku sme určili optimálny rozklad množiny Ω všetkých elementárnych javov a na základe tohoto pokusu našli podozrivú podmnožinu M množiny Ω. Potom sme pokračovali tak, ako keby sme položili $\Omega := M$ a použili rovnaký postup. Tak napríklad pri určovaní jedného žiaka z 32 bolo $\Omega = \{1, 2, \ldots, 32\}$ prvý pokus $\mathbf{P}_1 = \big\{\{1, 2\ldots, 16\}, \{17, 18, \ldots, 32\}\big\}$, ak jeho výsledkom bola prvá množina, pre druhý pokus bolo $\Omega = \{1, 2, \ldots, 16\}$ a $\mathbf{P}_2 = \big\{\{1, 2, \ldots, 8\}, \{9, 10, \ldots, 16\}\big\}$, v prípade, že druhý pokus poukázal na druhú množinu, pre tretí pokus definujeme $\Omega = \{9, 10, \ldots, 16\}$ a $\mathbf{P}_= \big\{\{9, 10, \ldots, 12\}, \{13, 14, \ldots, 16\}\big\}$, atď. Po vykonaní každého pokusu sa nám postup vetví na toľko možností, koľko výsledkov mal tento pokus. Ak by sme chceli presne popísať celý postup musíme popísať, čo máme robiť v každom vrchole príslušného stromu vetvenia.

Existuje ešte iný pohľad na túto problematiku. Každý pokus budeme robiť na základnom priestore

Ω tak, že definujeme postupnosť pokusov nasledovne"

$$\mathbf{P}_1 = \big\{\{1,2,\ldots,16\},\{17,18,\ldots,32\}\big\}$$
$$\mathbf{P}_2 = \big\{\{1,2,\ldots,8,17,18,\ldots 24\},\{9,10,\ldots,16,24,25,\ldots,32\}\big\}$$
$$\mathbf{P}_3 = \begin{cases}\big\{\{1,2,3,4,9,10,11,12,17,18,19,20,25,26,27,28\},\\ \{5,6,7,8,13,14,15,16,21,22,23,24,29,30,31,32\}\big\}\end{cases}$$
$$\mathbf{P}_4 = \begin{cases}\big\{\{1,2,5,6,9,10,13,14,17,18,21,22,25,26,29,30\},\\ \{3,4,7,8,11,12,15,16,19,20,23,24,27,28,31,32\}\big\}\end{cases}$$
$$\mathbf{P}_5 = \begin{cases}\big\{\{1,3,5,7,9,11,13,15,17,19,21,23,25,27,29,31\},\\ \{2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32\}\big\}\end{cases}$$

Ak vykonáme všetky pokusy $\mathbf{P}_1$ až $\mathbf{P}_5$, dostaneme z ich výsledkov jednoznačne, ktorý jav základného priestoru Ω nastal. Všimnime si, že v tomto prípade nezáleží na poradí vykonania jednotlivých pokusov.

V doterajších úlohách sme hľadali takú organizáciu pokusov, ako pomocou ich minimálneho počtu **zaručene** určiť, ktorý elementárny 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 sitáciu. Preto predpokladáme rovnakú pravdepodobnosť elementárnych javov základného pokusu. Tento predpoklad naviac 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.

Telefónne vedenie z miesta P do miesta Q je 100 metrov dlhé. Niekde medzi miestami P, Q sa vedenie prerušilo. Nech bod R leží v strede vedenia P,Q, pravdepodobnosť, že chyba vznikne v úseku P,R je päťkrát väčšia než pravdepodobnosť vzniku chyby v úseku R,Q. Vedenie môžeme merať tak, že ho v ľubovoľnom X mieste „napichneme" a zistíme, či je spojenie medzi miestami P a X. Treba určiť postup identifikácie úseku nie dlhšieho než 1 meter tak aby stredný počet meraní bol minimálny. Ako základný pokus budeme zase brať pokus

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

v ktorom však

$$P\big(\langle 0,1)\big) = P\big(\langle 1,2)\big) = \cdots = P\big(\langle 49,50)\big) = \frac{5}{300}$$
$$P\big(\langle 50,51)\big) = P\big(\langle 51,52)\big) = \cdots = P\big(\langle 99,100\rangle\big) = \frac{1}{300}$$

$$H(\mathbf{B}) = H\left(\underbrace{\frac{5}{300}\cdots\frac{5}{300}}_{50\text{krát}},\ \underbrace{\frac{1}{300}\cdots\frac{1}{300}}_{50\text{krát}}\right) = -50.\frac{5}{300}.\log_2(\frac{5}{300}) - 50.\frac{1}{300}.\log_2(\frac{1}{300}) \qquad (2.28)$$

1.371 4.9224

2.5 Podmienená entropia

Označme $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ nejaký pokus na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$. Predpokladajme, že nastal 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 $H(\mathbf{B}) = H(P(B_1), P(B_2), \ldots, P(B_m)$ 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 sú dané dva pokusy $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$. Podmienenou entropiu pokusu $\mathbf{B}$ za predpokladu, že nastal jav A_i (alebo len za podmienky A_i) je

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

Zrekapitulujme si v krátkosti predchádzajúcu úvahu. Zaujímame sa o výsledok pokusu $\mathbf{B}$, ktorý má entropiu $H(\mathbf{B})$. Nastal jav $\omega \in \Omega$; my sme však dostali správu, že $\omega \in A_i$ a táto správa zmenila entropiu pokusu $\mathbf{B}$ 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 jednozančne číslo $H(\mathbf{B}|A_i)$ – toto priradenie je diskrétnou náhodnou veličinou na Ω[3] . Stredá 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 $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$. Podmienenou entropiu pokusu $\mathbf{B}$ za predpokladu, vykonania pokusu $\mathbf{A}$ (alebo len za podmienky $\mathbf{A}$) je

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

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

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

[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 $\chi_{A_i}(\omega) = 0$.

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 $\mathbf{A}$, $\mathbf{B}$ 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 naprv $\mathbf{B}$ 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 prepokladu, 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 platí vzťah (2.14). 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

$$H(\mathbf{A} \wedge \mathbf{B}) = (q_{11}, q_{12}, \ldots q_{1,m},\ q_{21}, q_{22}, \ldots q_{2,m},\ \ldots, q_{n1}, q_{n2}, \ldots q_{n,m}) =$$

$$= H(p_1, p_2, \ldots, p_n) + \sum_{j=1}^{m} p_i.H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \ldots, \frac{q_{im}}{p_i}\right) =$$

$$= H\big(P(A_1), P(A_2), \ldots, P(A_n)\big) + \sum_{j=1}^{m} P(A_i).H\left(\frac{P(A_i \cap B_1)}{P(A_i)}, \frac{P(A_i \cap B_2)}{P(A_i)}, \ldots, \frac{P(A_i \cap B_m)}{P(A_i)}\right) =$$

$$H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

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 čím je entropia $H(\mathbf{A})$ pokusu $\mathbf{A}$ väčšia, tým menšia je podmienená entropia $H(\mathbf{B}|\mathbf{A})$.

Definícia 2.6. 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 $(\Omega, \mathcal{A}, P)$. Hovoríme, že pokusy $\mathbf{A}$, $\mathbf{B}$ sú štatisticky nezávislé (alebo len nezávislé), ak pre každé $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ sú A_i, B_j nezávislé javy.

2.6 Spoločná informácia pokusov

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

Definícia 2.7. Stredné množstvo informácie $I(\mathbf{A},\mathbf{B})$ o pokuse $\mathbf{B}$ v pokuse $\mathbf{A}$ 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 $\mathbf{B}$ obsiahnutá v pokuse $\mathbf{A}$ sa rovná informácii o pokuse $\mathbf{A}$ obsiahnutej v pokuse $\mathbf{B}$. Preto sa niekedy hodnote $I(\mathbf{A},\mathbf{B})$ hovorí aj spoločná informácia pokusov $\mathbf{A}$, $\mathbf{B}$.

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 disjunktný rozklad priestoru Ω je

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

Pretože zjednotenie na pravej strane posledného výrazu je disjunktné, je

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

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

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

□

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 $\ln x \leq x - 1$, ktorá platí pre všetky $x > 0$, pričom rovnosť nastáva práve vtedy, keď $x = 1$.

$$P(A_i \cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) = P(A_i \cap B_j).\ln(2).\ln\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) \leq$$
$$\leq P(A_i \cap B_j).\ln(2).\left[\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) - 1\right] = \ln(2).\left[P(A_i).P(B_j) - P(A_i \cap B_j)\right],$$

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.

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

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 okamžikoch odmerať – vzorkovať a pokiaľ je vzorkovacia frekvcencia dvojnásobná ako maximálna frekvencia signálu, stačia tieto diskrétne vzorky na plnohodnotnú rekonštrukciu pôvodného signálu. Môžeme teda predpokladať, že zdroj produkuje časových okamžikoch $t = t_1, t_2, t_3, \ldots$ 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 okamžiky $t = t_1, t_2, t_3, \ldots$ 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 okamžiky, $t = t_1, t_2, t_3, \ldots$, 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 okamžikmi 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, budme 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 včítane prázdnej postupnosti $e = \{\}$. 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:

$$1.\quad P(e) = 1 \tag{3.1}$$

$$2.\quad \sum_{(x_1,\ldots,x_n)\in X^n} P(x_1,\ldots,x_n) = 1 \tag{3.2}$$

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

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 okamžiku 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 okamžikov od spustenia zdroja. Podmienka (3.1) hovorí, že za 0 časových okamžikov vyšle zdroj prázdne slovo s pravdepodobnosťou rovnou 1, druhá podmienka hovorí, že za n časových okamžikov 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 $x_1, x_2, \ldots, x_n$ bola rovná 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 lyngvistickom a našom zmysle. V lyngvistickom 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 nášej 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 okamžikoch $n, n+1, \ldots, n+m-1$. Túto pravdepodobnosť vypočítame nasledovne:

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

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

$$P_i(x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_n) \quad \text{pre každé } i \text{ a každé } x_1, x_2 \ldots, x_n \in X^n \tag{3.5}$$

Označme X_i diskrétnu náhodnú premennú, ktorá bude popisovať vyslanie jedného znaku zo zdroja $(\mathcal{Z}, P)$ 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 $(\mathcal{Z}, P)$ je nezávislý ak pre ľubovoľné $i,\ j,\ n,\ m$ také, že $i + n \leq j$ platí

$$P\big([X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n] \cap [X_j = y_1, X_{j+1} = y_2, \ldots, X_{j+m-1} = y_m]\big) = \\ = P\big([X_i = x_1, X_{i+1} = x_2, \ldots, X_{i+n-1} = x_n]\big).P\big([X_j = y_1, X_{j+1} = y_2, \ldots, X_{j+m-1} = y_m]\big) \quad (3.6)$$

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 [Černý] 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 slovo „ÉHO␣CHLAPA“ (medzeru „␣“ považujeme tiež za symbol abecedy) bude mať nenulovú pravdepodobnosť, kým slovo „ÉMU␣CHLAPA“ sa vyskytuje s nulovou pravdepodobnosťou.

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.

Majme stacionárny zdroj $(\mathcal{Z}, P)$. 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 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 $H(\mathbf{D}) = H(\mathbf{B}) = H(p_1, p_2, \ldots, p_m)$ a $\mathbf{C}_2 = \mathbf{B} \wedge \mathbf{D}$.

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

Predpokladajme, že pre pokus

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

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

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

Označme

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

potom

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

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. Entropiu zdroja $\mathcal{Z}$ definujeme ako

$$H(\mathcal{Z}) = -\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{(x_1,\ldots,x_n)\in Z} P(x_1, x_2, \ldots, x_n).\log_2(P(x_1, x_2, \ldots, x_n)). \tag{3.7}$$

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

$$\begin{aligned}
&\sum_{(x_1,\ldots,x_n)\in Z} P(x_1, x_2, \ldots, x_n).\log_2(P(x_1, x_2, \ldots, x_n)) = \\
&= \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\big[\log_2 P(x_1) + \log_2 P(x_2) + \cdots + \log_2 P(x_n)\big] = \\
&= \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_1) + \\
&+ \sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_2) + \\
&+ \ldots\ldots\ldots\ldots\ldots\cdots + \\
&\sum_{(x_1,\ldots,x_n)\in Z} P(x_1).P(x_2),\ldots,P(x_n).\log_2 P(x_n) = \\
&= \sum_{x_1\in Z} P(x_1).\log_2 P(x_1) \cdot \underbrace{\sum_{(x_2,\ldots,x_n)\in Z} P(x_2).P(x_3),\ldots,P(x_n)}_{=1} + \cdots = \\
&= \sum_{x_1\in Z} P(x_1).\log_2 P(x_1) + \sum_{x_2\in Z} P(x_2).\log_2 P(x_2) + \cdots + \sum_{x_3\in Z} P(x_3).\log_2 P(x_3) = \\
&= n.\sum_{x\in Z} P(x).\log_2 P(x)
\end{aligned}$$

a preto pre stacionárny nezávislý zdroj $\mathcal{Z}$ platí $H(\mathcal{Z}) = -\sum_{x\in Z} P(x).\log_2 P(x)$.

Ak je zdroj len stacionárny, limita (3.7) vôbec nemusí existovať.

Veta 3.1. 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*

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

Dôkaz. Túto vetu uvádzame bez dôkazu.

Označme

$$E(n,\varepsilon) = \left\{x_1, \ldots x_n \in Z^n \mid \left|\frac{1}{n}.\log_2 P(x_1, \ldots x_n) + H(\mathcal{Z})\right| < \varepsilon\right\} \tag{3.9}$$

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:

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

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:

$$(1-\varepsilon).2^{n(H(\mathcal{Z})-\varepsilon)} < |E(n,\varepsilon)| < 2^{n(H(\mathcal{Z})+\varepsilon)} \tag{3.10}$$

Množina všetkých slov sĺžky n sa teda rozpadne na významnú množinu $E(n,\varepsilon)$ ktorá má príbližne $2^{n.H(\mathcal{Z})}$ 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. Naviac sa používajú aj interpunkčné znamienka (čiarka, dvojbodka, bodkočiarka, pomlčka, uvodzovky, 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 3. Počet všetkých 8-znakových slov abecedy Z je teda 40^8, $|E(8,\varepsilon)|$ odhadneme na $2^{8.3} = 2^{24}$.

$$\text{Je } \frac{2^{24}}{40^8} = \frac{2^{24}}{8^8.5^8} = \frac{1}{5^8} = \frac{1}{390626} = 2,56.10^{-6}.$$

Množina $E(8,\varepsilon)$ obsahuje menej ako 3 milióntiny celkového počtu 8-znakových slov.

3.3 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 kartézskym 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 okamžikov) a kde

$$P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) = P(a_1,a_2,\ldots,a_n).P(b_1,b_2,\ldots,b_n) \tag{3.11}$$

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

Veta 3.2. *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 prepíšeme nasledovne:

1. $P(e) = 1$ (3.12)
2. $$\sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) = 1 \tag{3.13}$$
3. $$\sum_{(p_1,q_1),\ldots,(p_m,q_m)\in(A\times B)^m} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n),(p_1,q_1),\ldots,(p_m,q_m) = \\ = P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big) \tag{3.14}$$

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

$$\sum_{(p_1,q_1),\ldots,(p_m,q_m)\in(A\times B)^m} P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n),(p_1,q_1),\ldots,(p_m,q_m) =$$
$$= \sum_{p_1p_2\ldots p_m\in A^m} \sum_{q_1q_2\ldots q_m\in B^m} P(a_1,a_2,\ldots,a_n,p_1,p_2,\ldots,p_m).P(b_1,b_2,\ldots,b_n,q_1,q_2,\ldots,p_m) =$$
$$= \sum_{p_1p_2\ldots p_m\in A^m} P(a_1,a_2,\ldots,a_n,p_1,p_2,\ldots,p_m) \sum_{q_1q_2\ldots q_m\in B^m} P(b_1,b_2,\ldots,b_n,q_1,q_2,\ldots,p_m) =$$
$$= P(a_1,a_2,\ldots,a_n)\ .\ P(b_1,b_2,\ldots,b_n) = P\big((a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\big)$$

Analogicky sa dokáže jdnoduchší druhý vzťah. □

Veta 3.3. *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*

$$H(\mathcal{Z}_1 \times \mathcal{Z}_2) = H(\mathcal{Z}_1) + H(\mathcal{Z}_2) \tag{3.15}$$

Dôkaz.

$$H(\mathcal{Z}_1 \times \mathcal{Z}_2) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{(a_1,b_1),\ldots,(a_n,b_n)\in(A\times B)^n} P\big((a_1,b_1),\ldots,(a_n,b_n)\big).\log_2 P\big((a_1,b_1),\ldots,(a_n,b_n)\big) =$$

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

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

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

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

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

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

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

□

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 3.3 a matematickej indukcie dostaneme nasledujúcu vetu:

Veta 3.4. *Nech $\mathcal{Z}$ je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}^n)$ zdroja $\mathcal{Z}^n$ platí*

$$H(\mathcal{Z}^n) = n.H(\mathcal{Z}) \tag{3.16}$$

Definícia 3.8. Nech $\mathcal{Z} = (A^*, P)$. Označme $\mathcal{Z}_{(k)} = \big((A^k)^*, P_{(k)}\big)$ 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

$$P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = P(a_{11}, a_{12}, \ldots, a_{1k}, a_{21}, a_{22}, \ldots, a_{2k}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nk}) \tag{3.17}$$

Informačný zdroj $\mathcal{Z}_{(k)}$ vznikne z informačného zdroja $\mathcal{Z}$ tak, že zo zdroja $\mathcal{Z}$ budeme odoberať každý k-ty okamžik 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 3.5. *Nech $\mathcal{Z}$ je informačný zdroj s entropiou $H(\mathcal{Z})$. Potom pre entropiu $H(\mathcal{Z}_{(k)})$ zdroja $\mathcal{Z}_{(k)}$ platí*

$$H(\mathcal{Z}_{(k)}) = k.H(\mathcal{Z}) \tag{3.18}$$

Dôkaz. Platí

$$\begin{aligned} H(\mathcal{Z}_{(k)}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{\mathbf{a}_1,\ldots,\mathbf{a}_n \in A^n} P_{(k)}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{a_{ij}\in A \text{ pre } 1\le i\le n,\ 1\le j\le k} P(a_{11}, a_{12}, \ldots, a_{1k}, a_{21}, a_{22}, \ldots, a_{2k}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nk}) = \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{x_1,x_2,\ldots,x_{n.k}\in A} P(x_1, x_2, \ldots, x_{n.k}) = \\ &= k. \left[\lim_{n\to\infty} \frac{1}{k.n} \sum_{x_1,x_2,\ldots,x_{n.k}\in A} P(x_1, x_2, \ldots, x_{n.k}) \right] = k.H(\mathcal{Z}) \end{aligned} \tag{3.19}$$

□

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.

3.4 Informačný zdroj ako súčin priestorov s mierou

Aj keď pomocou modelu z predchádzajúcej kapitoly dokážeme odvodiť všetky 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.

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

$$\Omega = \prod_{i=-\infty}^{\infty} Z \tag{3.20}$$

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 náhodnú funkciu X_i predpisom

$$X_i(\boldsymbol{\omega}) = \omega_i. \tag{3.21}$$

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

$$C_n(E_1, E_2, \dots, E_k) = \{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) \in E_1, X_{n+1}(\boldsymbol{\omega}) \in E_2, \dots, X_{n+k-1}(\boldsymbol{\omega}) \in E_k\} \tag{3.22}$$

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

$$EC_n(x_1, x_2, \dots, x_k) = \{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) = x_1, X_{n+1}(\boldsymbol{\omega}) = x_2, \dots, X_{n+k-1}(\boldsymbol{\omega}) = x_k\} \tag{3.23}$$

Všimnime si, že môžeme písať $T^1(\boldsymbol{\omega} = T(\boldsymbol{\omega}$, odobne

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

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

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 konečené prieniky a na konečené zjednotenia. Preto existuje najmenšia σ-algebra $\mathcal{F}$ podmnožín priestoru Ω obsahujúca $\mathcal{F}_0$.

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

$$T^{-1}A = \{\omega \mid T(\omega) \in A\} \qquad T(A) = \{T(\omega) |\ \omega \in A\} \tag{3.24}$$

Indukciou môžeme definovať $T^{-n}A$ takto: $T^{-1}A$ je definované v (3.24). 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

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

Hovoríme, že množina $B \in \mathcal{F}$ je T-invariantná, 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 3.6. *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 $B \in F$ je T -invariantná. Potom

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

Veta 3.7. 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 $\omega \in \Omega$ platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_A\left(T^i(\omega)\right) = P(A), \tag{3.26}$$

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

Definícia 3.10 a vety 3.6, 3.7 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

$$X_n(T(\boldsymbol{\omega})) = X_{n+1}(\boldsymbol{\omega}) \tag{3.27}$$

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

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

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

Zobrazenie T spolu s pravdepodobnostnou mierou P podstatne charakterizuje 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 3.8. *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 zachovúce zobrazenie.*

Dôkaz predchádzajúcej vety vyžaduje znalosť postupov teórie miery, preto ho neuvádzame. Pre prístup k modelovaniu zdrojov pomocou aparátu teórie miery je však táto veta typická tým, že v mnohých prípadoch stačí dokázať nejakú vlastnoť 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

$$P\big(EC_n(a_{i_1}, a_{i_2}, \ldots, a_{i_k})\big) = p_{i_1}.p_{i_2}, \ldots, p_{i_k}. \tag{3.28}$$

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

$$P\big(C_n(E_1, E_2, \ldots, E_k)\big) = P(E_1).P(E_2).\ldots.P(E_k). \tag{3.29}$$

Veta 3.9. *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 zachovajúce mieru.*

Teória miery teda zaručuje existenciu jedinej miery P na $\mathcal{F}$ splňujúcej (3.28). 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

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

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 (3.29) $\prod_{i=1}^{k} P(E_i).\prod_{j=1}^{l} P(F_j) = P(A).P(B)$. Ak $A,\ B$ sú cylindre, máme

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A).P(B) \tag{3.30}$$

Opäť si pomôžeme vetou z teórie miery.

Veta 3.10. *Nech $\mathcal{F}_0$ je algebra generujúca σ-algebru $\mathcal{F}$. Ak* (3.30) *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 $\Omega,\ \mathcal{F},\ T$ 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

$$P\{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) \in E_1, X_{n+1}(\boldsymbol{\omega}) \in E_2, \ldots, X_{n+k-1}(\boldsymbol{\omega}) \in E_k, \} \tag{3.31}$$

nezávisí na n, čo je ekvivaletné 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

$$P_k(x_1, x_2, \ldots, x_k) = P\{\boldsymbol{\omega} \mid X_n(\boldsymbol{\omega}) = x_1, X_{n+1}(\boldsymbol{\omega}) = x_2, \ldots, X_{n+k-1}(\boldsymbol{\omega}) = x_k\} \tag{3.32}$$

Pretože musí platiť (3.31) a pretože T zachováva mieru, musí byť

1. $P_k(x_1, x_2, \ldots, x_k) \geq 0$ (3.33)
2. $\sum_x P_{k+1}(x_1, x_2, \ldots, x_k, x) = P_k(x_1, x_2, \ldots, x_k)$ (3.34)
3. $\sum_x P_1(x) = 1$ (3.35)
4. $\sum_x P_{k+1}(x, x_1, x_2, \ldots, x_k) = P_k(x_1, x_2, \ldots, x_k)$ (3.36)

Naopak, ak máme systém funkcií $P_k(x, \ldots, x_k)$ splňujúcich (3.33), (3.34), (3.35) a (3.36), potom existuje jediná miera P na $\mathcal{F}$, ktorá sa zachováva pri zobrazení T a pre ktorú platí (3.32).

Príklad 3.3. Nech $\boldsymbol{\Pi} = (q_{ij})$ je stochastická matica typu $r \times r$, ktorej riadky a stĺpce zodpovedajú prvkom abecedy Z. 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

$$P_k(x_1, x_2, \ldots, x_k) = p_{x_1}.q_{x_1x_2}.q_{x_2x_3}.\ldots.q_{x_{k-1}x_k} \tag{3.37}$$

Ľahko overíme, že funkcie $P_k()$ definované v (3.37) spĺňajú (3.33), (3.34), (3.35) a pretože $\boldsymbol{p}.\boldsymbol{\Pi} = \boldsymbol{p}$, platí aj (3.36). Existuje teda miera P na $\mathcal{F}$, pre ktorú platí (3.32). 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

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

Nasledujúce štyri tvrdenia sú ekvivalentné

a) Zobrazenie T je ergodické.

b) Veličiny s_{ij} nezávisia od i.

c) Matica $\mathbf{\Pi}$ je neprivodima.

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 priblíženmím 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 predpokladť, 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.