Entropie, informace a kódování

Petr Kůrka

Obsah

1 Informace a entropie

1.1 Shannonova formule

Uvažujme hru, ve které si partner myslí prvek nějaké konečné abecedy A a máme tento prvek určit řadou otázek, na které lze odpovědět ano či ne. Má-li například abeceda $A = \{a, b, c, d\}$ 4 prvky, lze myšlený prvek zjistit dvěma otázkami. Ptáme se nejprve, zda prvek patří do množiny $\{a, b\}$. Pokud ano, ptáme se zda je to a. Pokud ne, ptáme se zda je to c. Alternativně lze tento princip vyjádřit kódováním. Sled odpovědí, které dostaneme, určují jednoznačně prvek dané abecedy. Existuje tedy kód, tj. prosté zobrazení $f : A \to B^2$ do binární abecedy $B = \{0, 1\}$ daný předpisem

x	a	b	c	d
$f(x)$	00	01	10	11

Má-li abeceda A 2^p prvků, lze každé její písmeno zjistit p binárními otázkami, neboli lze ho kódovat binárním slovem délky p. Říkáme že **entropie** abecedy A je $\mathcal{H}(A) = p = \log |A|$ bitů. Zde $\log = \log_2$ je logaritmus při základu 2 a **bit** je jednotka informace odpovídající písmenu binární abecedy.

Není-li počet prvků $|A|$ mocnina dvou, tj. platí-li $2^{p-1} < |A| \le 2^p$, potřebujeme pro kódování abecedy A také p bitů. Předpokládejme však, že místo jediného písmene máme určit celou zprávu v abecedě A, tj. slovo (posloupnost písmen) $u = u_0 u_1 u_2 \dots u_{n-1} \in A^*$. Kódujeme-li místo písmen celé bloky písmen, může být délka kódu kratší, než kdybychom kódovali každé písmeno zvlášť. Předpokládejme, že zprávu kódujeme po blocích délky m. Takových bloků je $|A|^m$, takže pro každý blok potřebujeme p_m bitů, kde $2^{p_m - 1} < |A|^m \le 2^{p_m}$. Odtud po logaritmování

$$\frac{p_m - 1}{m} < \log |A| \le \frac{p_m}{m} \implies \lim_{m \to \infty} \frac{p_m}{m} = \log |A|$$

Tento vztah je znám jako Heartleyho formule.

Definice 1 (Heartleyho formule) *Entropie konečné abecedy A je*

$$\mathcal{H}(A) = \log |A|.$$

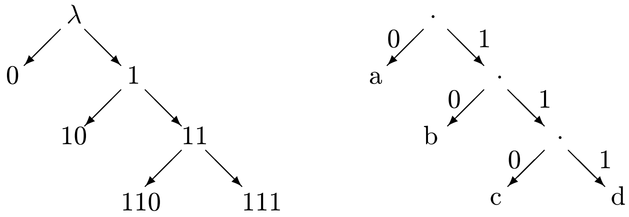
Jestliže se v kódované zprávě různá písmena vyskytují s různou pravděpodobností, lze délku kódu snížit tak, že písmenům s větší pravděpodobností přiřadíme kratší kódy a písmenům s menší pravděpodobností přiřadíme delší kódy. Pro danou abecedu A označme A^* množinu konečných slov a A^+ množinu neprázdných slov. Říkáme, že slovo $u \in A^*$ je **prefix** (počáteční úsek) slova $v \in A^*$, pokud existuje $w \in A^*$ takové že $v = uw$. To zahrnuje i případ $u = v$, kdy $w = \lambda$ je prázdné slovo nulové délky. Je-li u prefix v, píšeme $u \sqsubseteq_p v$.

Uvažujme abecedu $A = \{a, b, c, d\}$, jejíž písmena se vyskytují s pravděpodobnostmi $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$. Představme si, že místo písmene a se ve zprávě vyskytuje jedno z písmen a_0, a_1, a_2, a_3, která mezi sebou nedokážeme rozlišit, a podobně písmeno b zastupuje některé z písmen b_0, b_1. Místo abecedy A pak máme abecedu $A' = \{a_0, a_1, a_2, a_3, b_0, b_1, c, d\}$ s osmi prvky, které mají všechny stejnou pravděpodobnost $1/8$ a můžeme je kódovat binárním kódem $f' : A' \to B^3$

x	a_0	a_1	a_2	a_3	b_0	b_1	c	d
$f'(x)$	000	001	010	011	100	101	110	111

Přejdeme-li nyní zpět k abecedě A, vidíme že písmeno a je určeno společným prefixem 0 kódů písmen a_i a podobně b je určeno společným prefixem 10 kódů písmen b_i. Tak dostáváme kód $f : A \to B^*$

x	a	b	c	d
$f(x)$	0	10	110	111



Obrázek 1: Binární strom a prefixový kód

Přestože kódy písmen mají různou délku, určuje kód $f(u) = f(u_0)f(u_1)\dots f(u_{n-1})$ jednoznačně slovo $u = u_0 \cdots u_{n-1} \in A^*$. Kód f je totiž **prefixový**: jsou-li $x, y \in A$ různá písmena, není $f(x)$ prefix $f(y)$. Prefixový kód lze popsat strukturou binárního stromu. **Binární strom** je každá množina $T \subseteq B^*$, která pro každé $u \in T$ obsahuje také každý prefix $v \sqsubseteq_p u$. **List** binárního stromu je jeho prvek, který není vlastním prefixem žádného jiného prvku T. Například $T = \{\lambda, 0, 1, 10, 11, 110, 111\}$ je binární strom s listy $L(T) = \{0, 10, 110, 111\}$ (viz obrázek 1). Binární strom určuje orientovaný graf jehož vrcholy jsou prvky T a označené hrany jsou $u \xrightarrow{a} ua$, kde $a \in A$. Je-li T binární strom, pak každé prosté zobrazení $f : A \to L(T)$ je prefixový kód.

Je-li nyní $x = x_0 \dots x_{m-1}$ zpráva délky m v abecedě $A = \{a, b, c, d\}$ s pravděpodobnostním rozdělením $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$, je v ní přibližně $m/2$ písmen a s kódem délky 1, $m/4$ písmen b s kódem délky 2, atd. Očekávaná délka kódu je tedy

$$\frac{m}{2} \cdot 1 + \frac{m}{4} \cdot 2 + \frac{m}{8} \cdot 3 + \frac{m}{8} \cdot 3 = \frac{7m}{4}$$

To je méně než $2m$, což je délka kódu při standardním kódování slovy délky 2. Obecněji předpokládejme, že abeceda $A = \mathbb{Z}_k = \{0, 1, \dots, k-1\}$ má k prvků, a že pravděpodobnosti výskytu písmen jsou záporné mocniny dvou $P = (2^{-d_0}, \dots, 2^{-d_{k-1}})$. Stejným postupem jako v předcházejícím případě lze sestrojit prefixový kód $f : A \to B^*$ takový, že délka kódu písmene $a \in A$ je $|f(a)| = d_a = -\log P(a)$. Říkáme že **informační obsah** písmene a je $\mathfrak{I}(a) = -\log P(a)$. Délka kódu je pak

$$L(P, f) = \sum_{a \in A} 2^{-d_a} \cdot d_a = \sum_{a \in A} P(a) \cdot \mathfrak{I}(a) = \sum_{i \in A} P(a) \cdot \log \frac{1}{P(a)}.$$

Tuto hodnotu nazýváme **entropií** rozdělení P. Pro obecné pravděpodobnostní rozdělení dokážeme existenci kódu splňující podobný vztah ve Větě 54.

Definice 2 (Shannonova formule) Pravděpodobnostní rozdělení *nad abecedou A je vektor $P = (P(a))_{a \in A}$ nezáporných čísel, jejichž součet je 1. Pravděpodobnostní rozdělení tvoří* **simplex**

$$\Delta(A) = \{P \in [0, 1]^A : \sum_{a \in A} P(a) = 1\}$$

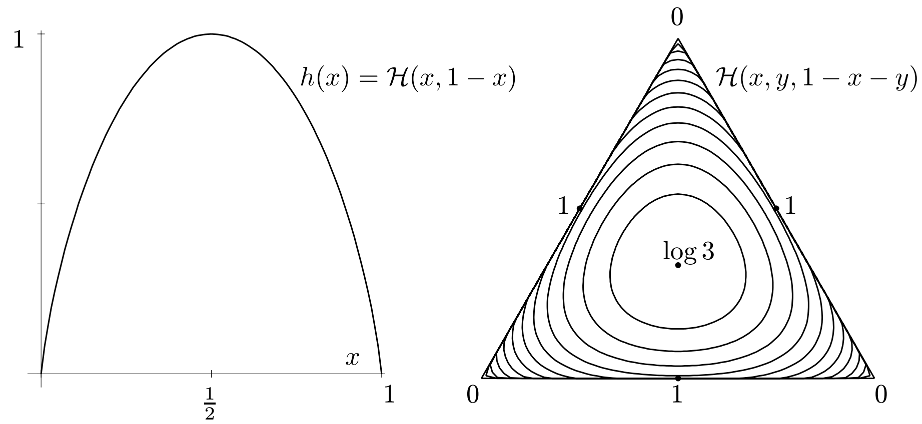
Informační obsah *písmene $a \in A$ a* **entropie** *rozdělení $P \in \Delta(A)$ je*

$$\mathfrak{I}_P(a) = \log \frac{1}{P(a)}, \quad \mathcal{H}(P) = -\sum_{a \in A} P(a) \cdot \log P(a)$$

V případě, že pravděpodobnost $P(a) = 0$ některého písmene je nulová, klademe $\mathfrak{I}_P(a) = \infty$ a $0 \cdot \log 0 = 0$, protože $\lim_{x \to 0^+} x \log(x) = 0$. Je-li rozdělení P rovnoměrné, tj. $P(a) = 1/|A|$, je $\mathfrak{I}_P(a) = \mathcal{H}(P) = \log|A|$. V tomto smyslu Shannonova formule zobecňuje Heartleyho formuli. Na obrázku 2 vlevo je entropie rozdělení $P = (x, 1-x)$ binární abecedy:

$$h(x) = \mathcal{H}(x, 1-x) = -x \cdot \log x - (1-x) \cdot \log(1-x)$$

Vidíme, že maximální entropii $\log 2 = 1$ má rovnoměrné rozdělení $P = (\frac{1}{2}, \frac{1}{2})$. Na obrázku 2 vpravo je entropie pravděpodobnostních rozdělení tříprvkové abecedy. Maximální entropii $\log 3 \approx 1.585$ má opět rovnoměrné rozdělení $P = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.



Obrázek 2: Entropie pravděpodobnostního rozdělení

Závislost délky kódu na frekvenci se uplatňuje i v přirozeném jazyce. Po vzniku pohyblivých obrázků před sto lety lidé nejprve chodili do kinematoskopu nebo do biografu. Po rozšíření této zábavy se začalo chodit do bia nebo do kina. Místo často se vyskytujících sousloví se používají zkratky. Dlouhé názvy jako například "Spojené království Velké Británie a Severního Irska" jsou ve zprávách většinou zastoupeny zkráceným názvem.

1.2 Kombinační čísla

Shannonovu formuli lze odvodit ještě jiným způsobem. Nechť abeceda $A = \{0, 1, \ldots, k-1\}$ má k prvků a nechť ve zprávě $u \in A^m$ délky m se vyskytuje právě m_i písmen i. Takové zprávy tvoří **kombinační množinu**

$$C(m_0, \ldots, m_{k-1}) = \{u \in A^m : \forall i < k, |u^{-1}(i)| = m_i\}, \text{ kde } m = m_0 + \cdots + m_{k-1}.$$

Počet takových zpráv je zobecněné **kombinační číslo**. Například pro binární abecedu je

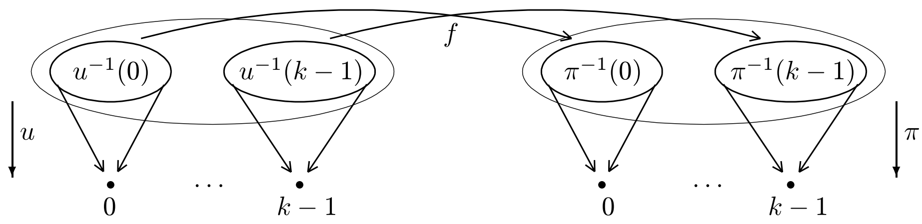
$$C(2,2) = \{0011, 0101, 0110, 1001, 1010, 1100\}, \quad |C(2,2)| = \frac{4!}{2! \cdot 2!} = 6$$

Tvrzení 1

$$|C(m_0, \ldots, m_{k-1})| = \binom{m_0 + \cdots + m_{k-1}}{m_0, \ldots, m_{k-1}} = \frac{(m_0 + \cdots + m_{k-1})!}{m_0! \cdots m_{k-1}!}$$

Důkaz: Položme $m = m_0 + \cdots + m_{k-1}$ a označme $\mathcal{P}(m)$ grupu permutací množiny $\mathbb{Z}_m$. Zvolme pevné $\pi \in C(m_0, \ldots, m_{k-1})$ a definujme zobrazení $F : \mathcal{P}(m) \to C(m_0, \ldots, m_{k-1})$ předpisem $F(f) = \pi f$ (viz obrázek 3). Pro každé $u \in C(m_0, \ldots, m_{k-1})$ platí

$$\begin{aligned} F^{-1}(u) &= \{f \in \mathcal{P}(m) : f[u^{-1}(0)] = \pi^{-1}(0), \ldots, f[u^{-1}(k-1)] = \pi^{-1}(k-1)\} \\ |F^{-1}(u)| &= m_0! \cdots m_{k-1}! \end{aligned}$$



Obrázek 3: Kombinační čísla

Odtud $|C(m_0, \dots, m_{k-1})| = |\mathcal{P}(m)|/|F^{-1}(u)| = m!/m_0! \cdots m_{k-1}!$ □

Velikost kombinační množiny odhadneme pomocí Stirlingovy formule $n! \approx (n/e)^n \sqrt{2\pi n}$.

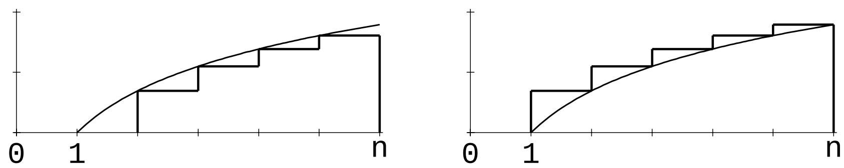
Lemma 2

$$\lim_{n\to\infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1, \quad \lim_{n\to\infty} \left(\frac{\log n!}{n} - \log \frac{n}{e} \right) = 0.$$

Důkaz: Stirlingova formule se dokazuje pomocí aproximace integrálu přirozeného logaritmu. Důkaz první formule vyžaduje dosti jemné aproximační metody (Wallisův součin). Ukážeme si důkaz druhé formule, jejíž důkaz je jednodušší (viz obrázek 4):

$$\begin{array}{rcl} \ln(2) + \cdots + \ln(n-1) < & \int_1^n \ln(x)\, dx & < \ln(2) + \cdots + \ln(n) \\ \ln(n!) - \ln(n) < & [x \ln(x) - x]_1^n & < \ln(n!) \\ n \ln(n) - n + 1 < & \ln(n!) & < (n+1)\ln(n) - n + 1 \end{array}$$

Po vydělení n dostáváme $\lim_{n\to\infty}(\ln(n!)/n - \ln(n) + 1) = 0$. Odtud převedením na dvojkový logaritmus $\lim_{n\to\infty}(\log(n!)/n - \log(n/e)) = 0$. □



Obrázek 4: Integrace logaritmu

Věta 3 *Nechť $m_n(0), \dots, m_n(k-1)$ jsou celočíselné posloupnosti a předpokládejme že pro $M_n = m_n(0) + \cdots + m_n(k-1)$ platí $\lim_{n\to\infty} M_n = \infty$, $\lim_{n\to\infty} \frac{m_n(i)}{M_n} = P(i)$. Pak*

$$\lim_{n\to\infty} \frac{\log |C(m_n(0), \dots, m_n(k-1))|}{M_n} = -\sum_{i=0}^{k-1} P(i) \cdot \log P(i) = \mathcal{H}(P)$$

Důkaz: Podle Stirlingova vzorce platí

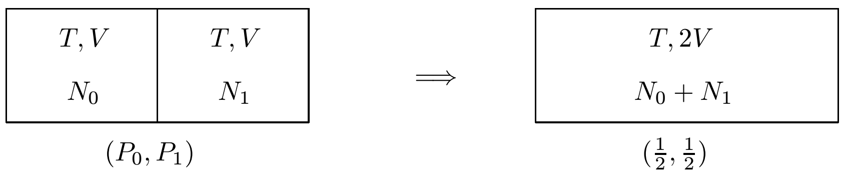
$$\begin{aligned} \lim_{n\to\infty} \frac{\log |C(m_n(0), \dots, m_n(k-1))|}{M_n} &= \lim_{n\to\infty} \frac{1}{M_n} \left(\log(M_n!) - \sum_{i<k} \log(m_n(i)!) \right) \\ &= \lim_{n\to\infty} \left(\frac{\log(M_n!)}{M_n} - \sum_{i<k} \frac{m_n(i)}{M_n} \cdot \frac{\log(m_n(i)!)}{m_n(i)} \right) \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \left(\log \frac{M_n}{e} - \sum_{i<k} P(i) \cdot \log \frac{m_n(i)}{e} \right) \\
&= \lim_{n\to\infty} \sum_{i<k} P(i) \cdot (\log(M_n) - \log(m_n(i))) \\
&= -\sum_{i<k} P(i) \cdot \log P(i). \ \square
\end{aligned}
$$

Pojem entropie v teorii informace souvisí s pojmem entropie v termodynamice. Termodynamické systémy jsou charakterizovány makroskopickými stavovými veličinami jako je tlak, teplota, objem, energie a také entropie. První termodynamický zákon říká, že energie izolovaného termodynamického systému zůstává konstantní. Druhý termodynamický zákon říká, že entropie izolovaného termodynamického systému nemůže klesat. Ve statistické termodynamice se rozlišuje makrostav systému od jeho mikrostavu, který je dán pozicemi a rychlostmi všech jeho částic (molekul). V kvantové mechanice danému makrostavu odpovídá velký ale konečný počet mikrostavů a entropie makrostavu se definuje (v souladu s Heartleyho formulí) jako logaritmus tohoto počtu mikrostavů (násobený Boltzmannovou konstantou). Pro entropii ideálního plynu se odvozuje přibližný vzorec

$$
S(T,V,N) = kN \left(\frac{3}{2} \ln T + \ln V - \ln N + C \right)
$$

Zde T je absolutní teplota, V je objem, N je počet molekul, k je Boltzmannova konstanta a C je plynová konstanta.



Obrázek 5: Růst entropie

Uvažujme nyní dvě nádoby o stejném objemu, ve kterých je různé množství téhož ideálního plynu při stejné teplotě a tedy různých tlacích. Stavy těchto systémů jsou tedy (T, V, N_0), (T, V, N_1) (obrázek 5) a plyn je mezi těmito nádobami rozdělen pravděpodobnostním vektorem

$$
P = (P_0, P_1) = \left(\frac{N_0}{N_0 + N_1}, \frac{N_1}{N_0 + N_1} \right)
$$

Spojíme-li tyto dva systémy do jednoho, bude výsledný stav $(T, 2V, N_0 + N_1)$. Změna entropie po tomto spojení je

$$
\begin{aligned}
S - S_0 - S_1 &= k(N_0 + N_1)(\tfrac{3}{2} \ln T + \ln 2 + \ln V - \ln(N_0 + N_1) + C) \\
&- kN_0(\tfrac{3}{2} \ln T + \ln V - \ln N_0 + C) - kN_1(\tfrac{3}{2} \ln T + \ln V - \ln N_1 + C) \\
&= k(N_0 + N_1)(\ln 2 - \ln(N_0 + N_1)) + kN_0 \ln N_0 + kN_1 \ln N_1 \\
&= k(N_0 + N_1)(\ln 2 + P_0 \ln P_0 + P_1 \ln P_1) \\
&= k(N_0 + N_1) \left(\mathcal{H}(\tfrac{1}{2}, \tfrac{1}{2}) - \mathcal{H}(P_0, P_1) \right)
\end{aligned}
$$

Změna entropie je tedy rozdíl (informačně-teoretických) entropií koncového rozdělení $(\frac{1}{2}, \frac{1}{2})$ a počátečního rozdělení (P_0, P_1) násobený celkovým množstvím látky.

1.3 Divergence entropie

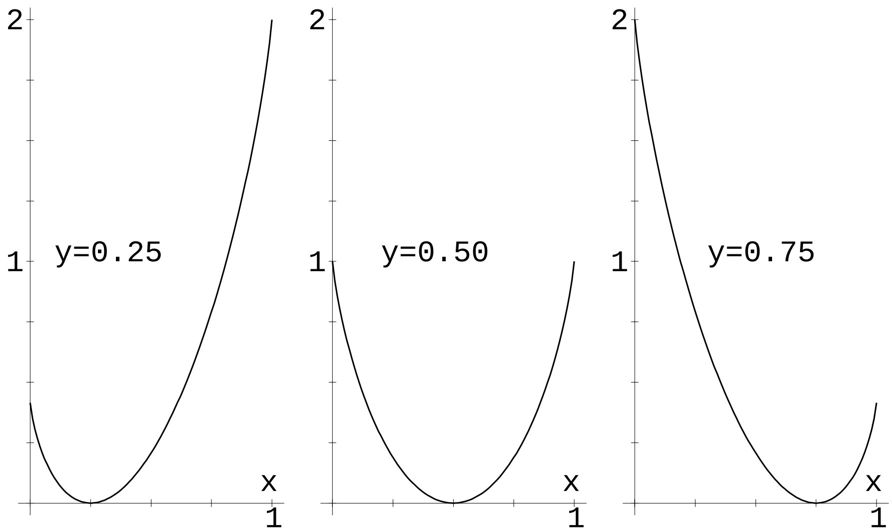
Uvažujme nyní situaci, kdy rozdělení $P \in \Delta(A)$ neznáme, a předpokládáme, že písmena zprávy mají rozdělení $Q \in \Delta(A)$. Kódujeme-li zprávu s tímto předpokladem, potřebujeme $-\log Q(a)$ bitů na písmeno a, takže celkové množství informace (délka kódu) je

$$\sum_{a \in A} P(a) \cdot \log \frac{1}{Q(a)}.$$

Ukážeme si (Tvrzení 6), že tato hodnota je vždy větší nebo rovna entropii $\mathcal{H}(P)$. Rozdíl mezi těmito dvěma hodnotami nazýváme divergenci entropie.

Definice 3 **Divergence entropie** *rozdělení $P \in \Delta(A)$ vzhledem k rozdělení $Q \in \Delta(A)$ (nad stejnou abecedou A) je definována vzorcem*

$$\mathcal{D}(P||Q) = \sum_{a \in A} P(a) \cdot \log \frac{1}{Q(a)} - \mathcal{H}(P) = \sum_{a \in A} P(a) \cdot \log \frac{P(a)}{Q(a)}.$$



Obrázek 6: Divergence entropie

Na obrázku 6 je divergence entropie pravděpodobnostních rozdělení dvouprvkové abecedy

$$d(x, y) = x \cdot \log \frac{x}{y} + (1-x) \cdot \log \frac{1-x}{1-y}.$$

Platí $d(x, \frac{1}{2}) = 1 - h(x)$, $d(0, y) = -\log(1-y)$, $d(1, y) = -\log y$. Vlastnosti funkce entropie a divergence entropie odvozujeme z vlastností funkce přirozeného logaritmu.

Definice 4 *Reálná funkce $f : I \to \mathbb{R}$ je* **konvexní** *na intervalu I, pokud její graf leží pod každou její sečnou, tj, pokud pokud pro každé $x, y \in I$ a každé $a \in (0, 1)$ platí*

$$f(ax + (1-a)y) \leq a \cdot f(x) + (1-a) \cdot f(y)$$

Funkce f je **striktně konvexní**, *pokud pro každé $x \neq y \in I$ platí ostrá nerovnost*

$$f(ax + (1-a)y) < a \cdot f(x) + (1-a) \cdot f(y)$$

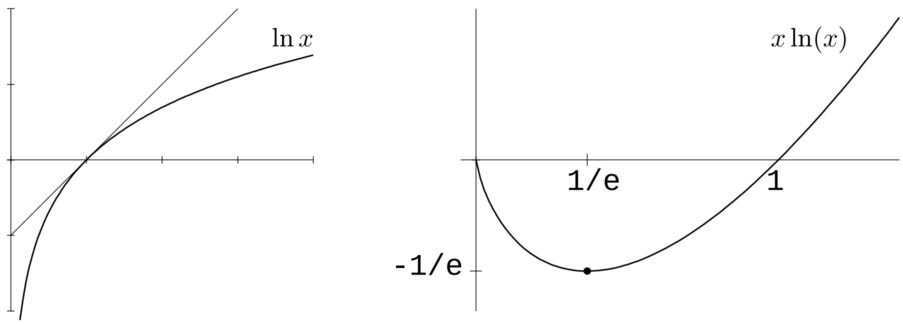
Funkce f je **(striktně) konkávní**, *je-li funkce $-f$ (striktně) konvexní.*

Tvrzení 4 (Jensenova nerovnost) *Nechť $f : I \to \mathbb{R}$ je konvexní funkce, $x_1, \ldots, x_n \in I$ a nechť $t_1, \ldots, t_n$ jsou nezáporná čísla, jejichž součet je 1. Pak*

$$f\left(\sum_{i\leq n} t_i x_i\right) \leq \sum_{i=1}^{n} t_i \cdot f(x_i)$$

Důkaz: Podle předpokladu tvrzení platí pro $n = 2$. Předpokládejme, že tvrzení platí pro $n-1$ a dokažme ho pro n:

$$\begin{aligned} f\left(\sum_{i\leq n} t_i x_i\right) &= f\left((1-t_n)\frac{\sum_{i<n} t_i x_i}{1-t_n} + t_n x_n\right) \leq (1-t_n) \cdot f\left(\frac{\sum_{i<n} t_i x_i}{1-t_n}\right) + t_n \cdot f(x_n) \\ &\leq (1-t_n)\sum_{i<n}\frac{t_i}{1-t_n} f(x_i) + t_n \cdot f(x_n) = \sum_{i=1}^{n} t_i \cdot f(x_i). \ \square \end{aligned}$$



Obrázek 7: Funkce $\ln(x)$ a $x\ln(x)$

Je-li $f''(x) > 0$ na otevřeném intervalu I, pak f je konvexní na I. Například funkce $\ln(x)$ je konkávní na $(0, \infty)$ a funkce $x \cdot \ln(x)$ je konvexní na $(0, \infty)$. Její limita v nule je nulová a minimum $-1/e$ nabývá v bodě $1/e$ (viz obrázek 7). Z konkávnosti logaritmu plyne nerovnost, kterou budeme opakovaně používat:

Lemma 5 *Pro každé kladé x platí $\ln(x) \leq x - 1$ a rovnost nastává pouze pro $x = 1$.*

Tvrzení 6 *Nechť $P, Q \in \Delta(A)$ jsou pravděpodobnostní rozdělení nad stejnou abecedou A. Pak $\mathcal{D}(P||Q) \geq 0$ a rovnost nastává právě když $P = Q$.*

Důkaz: Z Tvrzení 5 plyne

$$\sum_{a\in A} P(a) \cdot \ln\frac{P(a)}{Q(a)} = -\sum_{a\in A} P(a) \cdot \ln\frac{Q(a)}{P(a)} \geq -\sum_{a\in A} P(a) \cdot \left(\frac{Q(a)}{P(a)} - 1\right) = 0$$

Odtud $\mathcal{D}(P||Q) = \sum_{a\in A} P(a) \cdot \ln\frac{P(a)}{Q(a)}/\ln 2 \geq 0$. $\square$

Speciálně pro rovnoměrné rozdělení $Q = (\frac{1}{n}, \ldots, \frac{1}{n})$ je entropie maximální. Pro každé $P \in \Delta(\mathbb{Z}_n)$ je

$$\mathcal{H}(P) \leq \sum_{i=0}^{n-1} P(i) \cdot \log n = \log n = \mathcal{H}(Q).$$

Vlastnost konvexity lze uvažovat nejen pro reálné funkce ale i pro funkci entropie na simplexu $\Delta(A)$. Pro $P, Q \in \Delta(A)$ a $0 \le t \le 1$, položme $R(a) = tP(a) + (1-t)Q(a)$. Pak $R \in \Delta(A)$ je rozdělení, které nazýváme konvexní kombinaci rozdělení P, Q. Obecněji jsou-li $P_1, \dots, P_n \in \Delta(A)$ a $t_1, \dots, t_n$ nezáporná čísla jejichž součet je 1, je **konvexní kombinace** $Q = \sum_{i=1}^n t_i P_i$ (tj. $Q(a) = \sum_{i=1}^n t_i P_i(a)$ pro každé $a \in A$) pravděpodobnostní rozdělení na A.

Tvrzení 7 *Entropie je konkávní funkce na $\Delta(A)$, tj.*

$$\mathcal{H}\left(\sum_{i=1}^n t_i P_i\right) \ge \sum_{i=1}^n t_i \cdot \mathcal{H}(P_i).$$

Důkaz: Pro funkci $g(x) = -x \ln x$ platí $g'(x) = -\ln x - 1$, $g''(x) = -1/x < 0$, takže g je konkávní. To znamená že pro každá kladná $x_1, \dots, x_n$ a jejich konvexní kombinaci platí $g(t_1 x_1 + \cdots t_n x_n) \ge t_1 \cdot g(x_1) + \cdots t_n \cdot g(x_n)$. Odtud

$$\mathcal{H}\left(\sum_i t_i P_i\right) = \sum_{a \in A} g\left(\sum_i t_i P_i(a)\right) \ge \sum_{a \in A} \sum_i t_i \cdot g(P_i(a)) = \sum_i t_i \cdot \mathcal{H}(P_i) \ \Box$$

Relativní entropie $\mathcal{D}(P||Q)$ je naopak konvexní funkce (obou rozdělení P a Q). Nejprve si dokážeme pomocnou nerovnost.

Lemma 8 *Nechť $a_1 \dots, a_n, b_1, \dots, b_n$ jsou nezáporná reálná čísla. Pak*

$$\sum_{i=1}^n a_i \cdot \log \frac{a_i}{b_i} \ge \left(\sum_{i=1}^n a_i\right) \cdot \log\left(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\right)$$

Důkaz: Položme $t_i = b_i / \sum_{j=1}^n b_j$, $x_i = a_i / b_i$. Funkce $f(x) = x \log(x)$ je striktně konvexní. takže

$$\sum_{i=1}^n \frac{a_i}{\sum_{j=1}^n b_j} \log \frac{a_i}{b_i} = \sum_{i=1}^n t_i f(x_i) \ge f(\sum_{i=1}^n t_i x_i) = \sum_{i=1}^n \frac{a_i}{\sum_j b_j} \cdot \log\left(\sum_{i=1}^n \frac{a_i}{\sum_j b_j}\right)$$

a to už je Jensenova nerovnost. $\Box$

Tvrzení 9 *$\mathcal{D}(P||Q)$ je konvexní funkce na $\Delta(A) \times \Delta(A)$, tj.*

$$\mathcal{D}\left(\sum_i t_i P_i \middle\| \sum_i t_i Q_i\right) \le \sum_i t_i \cdot \mathcal{D}(P_i||Q_i)$$

Důkaz: Pro každé $a \in A$ platí podle Tvrzení 8

$$\left(\sum_i t_i P_i(a)\right) \log \frac{\sum_i t_i P_i(a)}{\sum_i t_i Q_i(a)} \le \sum_i t_i P_i(a) \log \frac{t_i P_i(a)}{t_i Q_i(a)}$$

Sečteme-li tyto nerovnosti pro všechna $a \in A$, dostaneme požadovanou nerovnost. $\Box$

1.4 Spočetná pravděpodobnostní rozdělení

Pravděpodobnostní rozdělení existují i na spočetných abecedách. Střední hodnotu a entropii spočetného pravděpodobnostního rozdělení $P = (P(n))_{n \in \mathbb{N}} \in \Delta(\mathbb{N})$ definujeme vzorcem

$$\mathbb{E}(P) = \sum_{i \in \mathbb{N}} i \cdot P(i), \quad \mathcal{H}(P) = -\sum_{i \in \mathbb{N}} P(i) \cdot \log P(i).$$

Příklad 1 *Entropie spočetného rozdělení může být nekonečná.*

Položme

$$a = \sum_{n=3}^{\infty} \frac{1}{n \ln^2 n} < \int_2^{\infty} \frac{dx}{x \ln^2 x} = \left[-\frac{1}{\ln x} \right]_2^{\infty} = \frac{1}{\ln 2}$$

Pro $P(n) = 1/an \ln^2 n$, kde $n \geq 3$ je

$$\mathcal{H}(P) = \frac{1}{\ln 2} \sum_{n=3}^{\infty} \frac{\ln a + \ln n + 2 \ln \ln n}{an \ln^2 n} > \frac{1}{\ln 2} \int_3^{\infty} \frac{dx}{ax \ln x} = \frac{1}{a \ln 2} [\ln \ln x]_3^{\infty} = \infty$$

Střední hodnota tohoto rozdělení je také nekonečná:

$$\mathbb{E}(P) = \sum_{n=3}^{\infty} \frac{1}{a \ln^2 n} > \sum_{n=3}^{\infty} \frac{1}{an} = \infty.$$

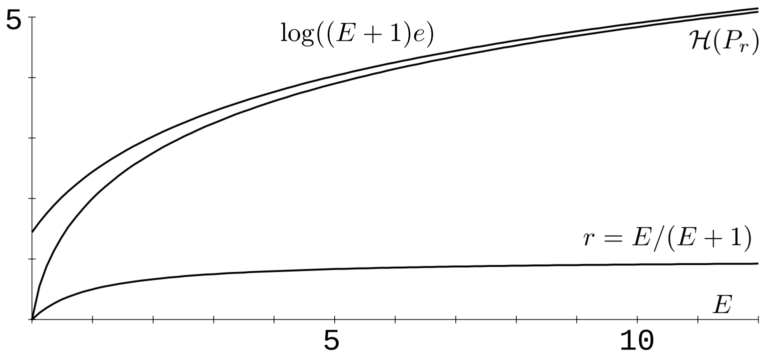
Příklad 2 *Pro geometrické rozdělení $P_r(n) = (1-r)r^n$, kde $0 < r < 1$, je*

$$\mathbb{E}(P_r) = \frac{r}{1-r}, \quad \mathcal{H}(P_r) = \log \frac{1}{1-r} + \frac{r}{1-r} \log \frac{1}{r} = \frac{\mathcal{H}(r, 1-r)}{1-r}.$$

Důkaz:

$$\begin{aligned}
\mathbb{E}(P_r) &= (1-r) \sum_{n=0}^{\infty} n r^n = (1-r) \left(\sum_{n=1}^{\infty} r^n + \sum_{n=2}^{\infty} r^n + \cdots \right) = r + r^2 + \cdots = \frac{r}{1-r} \\
\mathcal{H}(P_r) &= -\sum_{n=0}^{\infty} (1-r) r^n (\log(1-r) + n \log r) = -\log(1-r) - \frac{r}{1-r} \log r. \quad \square
\end{aligned}$$

Speciálně pro $r = \frac{1}{2}$ je $P_{\frac{1}{2}}(n) = 2^{-n-1}$, $\mathbb{E}(P_r) = 1$ a $\mathcal{H}(P_r) = 2$. Závislost entropie na střední hodnotě je na obrázku 8. Pro velká E, kdy r je blízko jedné, je $\mathcal{H}(P_r) \approx \log((\mathbb{E}(P_r) + 1)e)$.



Obrázek 8: Entropie geometrického rozdělení

Věta 10 *Je-li $Q \in \Delta(\mathbb{N})$ pravděpodobnostní rozdělení se střední hodnotou $E = \mathbb{E}(Q) < \infty$ a $P_r \in \Delta(\mathbb{N})$ geometrické rozdělení se stejnou střední hodnotou $\mathbb{E}(P_r) = E$, pak*

$$\mathcal{H}(Q) = \mathcal{H}(P_r) - \mathcal{D}(Q || P_r) \leq \mathcal{H}(P_r) \leq \log((E + 1)e).$$

Důkaz: Střední hodnota je $E = \mathbb{E}(Q) = \sum_{n=0}^{\infty} n \cdot Q(n)$. Pro $r = E/(E+1)$ je $P_r(n) = (1-r)r^n$ geometrické rozdělení se střední hodnotou E a platí

$$\begin{aligned}\sum_{n\in\mathbb{N}} Q(n)\cdot\log P_r(n) &= \sum_{n\in\mathbb{N}} Q(n)\cdot(\log(1-r)+n\cdot\log r) = \log(1-r)+E\cdot\log r\\ &= \sum_{n\in\mathbb{N}} P_r(n)\cdot(\log(1-r)+n\cdot\log r) = \sum_{n\in N} P_r(n)\log P_r(n)\end{aligned}$$

takže

$$\mathcal{H}(Q) = -\sum_{n\in\mathbb{N}} Q(n)\left(\log P_r(n) + \log\frac{Q(n)}{P_r(n)}\right) = \mathcal{H}(P_r) - \mathcal{D}(Q||P_r).$$

Po dosazení $r = E/(E+1)$ dostáváme

$$\begin{aligned}\mathcal{H}(P_r) &= \log(E+1) + E\cdot\log\frac{E+1}{E} \leq \log(E+1) + E\cdot\left(\frac{E+1}{E}-1\right)/\ln 2\\ &= \log(E+1)+\log e. \ \square\end{aligned}$$

Věta 10 je speciální případ principu, který se používá ve statistické mechanice, kde se hledá maximální entropie rozdělení, které splňují jistá omezení. Uvažujme matici nezáporných čísel $M = (M_{ij})_{i\in\mathbb{N},j=0,\ldots,r}$, kde $M_{i0} = 1$. Pro kladná čísla $\alpha_0 = 1, \alpha_1, \ldots, \alpha_r$ položme

$$W(\alpha_1,\ldots,\alpha_r) = \{Q\in\mathbb{R}_+^{\mathbb{N}} :\ \forall j\leq r, \sum_{i\in\mathbb{N}} Q(i)M_{ij} = \alpha_j\}$$

Pro $j = 0$ dostáváme $\sum_{i\in\mathbb{N}} Q(i) = 1$, takže každé $Q \in W(\alpha_1,\ldots,\alpha_r)$ je pravděpodobnostní rozdělení.

Tvrzení 11 *Předpokládejme, že existují $\lambda_0, \lambda_1, \ldots, \lambda_r$, taková že pro $P(i) = 2^{-\sum_{j=0}^{\infty} M_{ij}\lambda_j}$ platí $P \in W(\alpha_1,\ldots,\alpha_r)$. Pak pro každé $Q \in W(\alpha_1,\ldots,\alpha_r)$ platí*

$$\mathcal{H}(Q) = \mathcal{H}(P) - \mathcal{D}(Q||P) \leq \mathcal{H}(P).$$

Důkaz:

$$\begin{aligned}\mathcal{H}(Q) &= -\sum_{i\in\mathbb{N}} Q(i)\left(\log P(i) + \log\frac{Q(i)}{P(i)}\right) = \sum_{i\in\mathbb{N}} Q(i)\sum_{j=0}^{r} M_{ij}\lambda_j - \mathcal{D}(Q||P)\\ &= \sum_{j=0}^{r}\alpha_j\lambda_j - \mathcal{D}(Q||P) = \sum_{i\in\mathbb{N}} P(i)\sum_{j=1}^{r} M_{ij}\lambda_j - \mathcal{D}(Q||P)\\ &= \mathcal{H}(P) - \mathcal{D}(Q||P)\ \square\end{aligned}$$

1.5 Entropie náhodných veličin

Teorie informace se systematicky rozvíjí v rámci teorie pravděpodobnosti. Zavádí se pojem informačního obsahu pravděpodobnostního jevu a náhodné veličiny. Entropie náhodné veličiny je střední hodnota jejího informačního obsahu. Teorie informace pojednává zejména o informačních vztazích mezi dvěma a více náhodnými veličinami.

Pravděpodobnostní prostor je trojice $(\Omega, \mathcal{A}, \mathbb{P})$, kde Ω je množina elementárních jevů, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ je σ-algebra a $\mathbb{P} : \Omega \to [0,1]$ je pravděpodobnostní míra na Ω. **Náhodná veličina** je měřitelné zobrazení $X : \Omega \to A$, kde A je konečná nebo spočetná abeceda. To znamená, že vzor každého písmene $a \in A$

$$X^{-1}(a) = \{\omega\in\Omega :\ X(\omega) = a\} = [X = a] \in \mathcal{A}$$

je měřitelná množina, protože na abecedě A implicitně předpokládáme σ-algebru $\mathcal{P}(A)$ všech jejích podmnožin. Rozdělení $P_X \in \Delta(A)$ náhodné veličiny X je dáno vzorcem $P_X(a) = \mathbb{P}[X = a] = \mathbb{P}(X^{-1}(a))$. Říkáme, že náhodná veličina $X : \Omega \to A$ je **reálná**, pokud $A \subset \mathbb{R}$ je (nejvýše spočetná) množina reálných čísel. Například charakteristická funkce $\chi_M : \Omega \to \{0, 1\}$ náhodného jevu $M \in \mathcal{A}$ je definována vzorcem

$$\chi_M(\omega) = \begin{cases} 0 & \text{pro} \quad \omega \in \Omega \setminus M \\ 1 & \text{pro} \quad \omega \in M \end{cases}$$

Střední hodnota a rozptyl reálné náhodné veličiny $X : \Omega \to A \subset \mathbb{R}$ je

$$\mathbb{E}(X) = \sum_{a \in A} P_X(a) \cdot X(a), \quad \mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Například střední hodnota a rozptyl charakteristické funkce je $\mathbb{E}(\chi_M) = \mathbb{P}(M)$, $\mathbb{V}(\chi_M) = \mathbb{P}(M) \cdot (1 - \mathbb{P}(M))$.

Definice 5

(1) **Informační obsah** *náhodného jevu $M \in \mathcal{A}$ je $\mathfrak{I}(M) = -\log \mathbb{P}(M)$.*

(2) **Informační obsah** *náhodné veličiny $X : \Omega \to A$ je náhodná veličina $\mathfrak{I}_X : \Omega \to [0, \infty]$ daná předpisem $\mathfrak{I}_X(\omega) = -\log P_X(X(\omega))$, tj.*

$$\mathfrak{I}_X = -\sum_{a \in A} \log P_X(a) \cdot \chi_{X^{-1}(a)}$$

(3) **Entropie náhodné veličiny** *X je*

$$\mathcal{H}(X) = \mathbb{E}(\mathfrak{I}_X) = \mathcal{H}(P_X) = -\sum_{a \in A} P_X(a) \cdot \log P_X(a).$$

Informační obsah náhodné veličiny X je tedy složení zobrazení

$$\mathfrak{I}_X : \Omega \xrightarrow{X} A \xrightarrow{P_X} [0, 1] \xrightarrow{-\log} [0, \infty]$$

Jsou-li $M, N \in \mathcal{A}$ nezávislé jevy, tj. je-li $\mathbb{P}(M \cap N) = \mathbb{P}(M) \cdot \mathbb{P}(N)$, je

$$\mathfrak{I}(M \cap N) = \mathfrak{I}(M) + \mathfrak{I}(N).$$

Uvažujme náhodné veličiny $X : \Omega \to A$ a $Y : \Omega \to B$. Jejich dvojice je náhodná veličina $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to A \times B$. Rozdělení těchto náhodných veličin je

$$\begin{aligned} P_{X,Y}(a, b) &= \mathbb{P}[X = a, Y = b], \\ P_X(a) &= \sum_{b \in B} P_{XY}(a, b) = \mathbb{P}[X = a], \\ P_Y(b) &= \sum_{a \in A} P_{XY}(a, b) = \mathbb{P}[Y = b]. \end{aligned}$$

Náhodné veličiny X a Y jsou nezávislé právě když $P_{XY}(a, b) = P_X(a) \cdot P_Y(b)$.

Tvrzení 12 *Jsou-li X, Y nezávislé náhodné veličiny, pak $\mathfrak{I}_{X,Y} = \mathfrak{I}_X + \mathfrak{I}_Y$.*

Důkaz:

$$\begin{aligned} \mathfrak{I}_{X,Y}(\omega) &= -\log P_{X,Y}(X(\omega), Y(\omega)) = -\log(P_X(X(\omega)) \cdot P_Y(Y(\omega))) \\ &= \mathfrak{I}_X(\omega) + \mathfrak{I}_Y(\omega). \ \square \end{aligned}$$

Jsou-li X a Y závislé, není mezi $\mathfrak{I}_{X,Y}$ a $\mathfrak{I}_X + \mathfrak{I}_Y$ žádný vztah.

Příklad 3 *Uvažujme pravděpodobnostní rozdělení na abecedách $A = \{a_0, a_1\}$, $B = \{b_0, b_1\}$:*

$$P_{XY}: \begin{array}{c|cc} & b_0 & b_1 \\ \hline a_0 & 1/3 & 0 \\ a_1 & 1/2 & 1/6 \end{array}, \quad \mathfrak{I}_{X,Y}: \begin{array}{c|cc} & b_0 & b_1 \\ \hline a_0 & 1.58 & \infty \\ a_1 & 1.00 & 2.58 \end{array}, \quad \mathfrak{I}_X + \mathfrak{I}_Y: \begin{array}{c|cc} & b_0 & b_1 \\ \hline a_0 & 1.58 + 0.26 & 1.58 + 2.58 \\ a_1 & 0.58 + 0.26 & 0.58 + 2.58 \end{array}$$

Je totiž $P_X = [\frac{1}{3}, \frac{2}{3}]^T$ (transpozice), $P_Y = [\frac{5}{6}, \frac{1}{6}]$, $\mathfrak{I}_X = [1.58, 0.58]^T$, $\mathfrak{I}_Y = [0.26, 2.58]$. Pro entropii, tj. střední hodnotu informačního obsahu však platí nerovnost:

Tvrzení 13 *Pro náhodné veličiny X, Y platí*

$$\max\{\mathcal{H}(X), \mathcal{H}(Y)\} \leq \mathcal{H}(X, Y) \leq \mathcal{H}(X) + \mathcal{H}(Y),$$

a $\mathcal{H}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y)$ platí právě když X a Y jsou nezávislé.

Důkaz: Pro rozdělení $Q(a, b) = P_X(a) \cdot P_Y(b)$ na $A \times B$ platí podle Tvrzení 6

$$\begin{aligned} \mathcal{H}(X, Y) &= -\sum_{a \in A} \sum_{b \in B} P_{X,Y}(a, b) \cdot \log P_{XY}(a, b) \\ &\leq -\sum_{a \in A} \sum_{b \in B} P_{XY}(a, b) \cdot (\log P_X(a) \cdot P_Y(b)) \\ &= -\sum_{a \in A} \sum_{b \in B} P_{XY}(a, b) \cdot \log P_X(a) - \sum_{a \in A} \sum_{b \in B} P_{XY}(a, b) \cdot \log P_Y(b) \\ &= \mathcal{H}(X) + \mathcal{H}(Y). \end{aligned}$$

Rovnost zde platí právě když $P_{XY}(a, b) = P_X(a) \cdot P_Y(b)$, tj. když X, Y jsou nezávislé. Druhá část tvrzení plyne z nerovnosti $P_{XY}(a, b) \leq P_X(a)$:

$$\mathcal{H}(X, Y) = -\sum_{a \in A} \sum_{b \in B} P_{XY}(a, b) \cdot \log P_{XY}(a, b) \geq -\sum_{a \in A} \sum_{b \in B} P_{XY}(a, b) \cdot \log P_X(a) = \mathcal{H}(X).$$

a obdobně $\mathcal{H}(X, Y) \geq \mathcal{H}(Y)$, takže $\mathcal{H}(X, Y) \geq \max\{\mathcal{H}(X), \mathcal{H}(Y)\}$. $\square$

Definice 6 **Vzájemná informace** *dvou náhodných proměnných je*

$$\mathcal{I}(X : Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y).$$

Podle Tvrzení 13 je $0 \leq \mathcal{I}(X : Y) \leq \min\{\mathcal{H}(X), \mathcal{H}(Y)\}$ a $\mathcal{I}(X : Y) = 0$ platí právě když X a Y jsou nezávislé.

1.6 Podmíněná informace a entropie

Je-li pravděpodobnostní rozdělení P_X kladné, existují podmíněné pravděpodobnosti

$$P_{Y|X}(a, b) = \mathbb{P}[Y = b | X = a] = P_{XY}(a, b) / P_X(a).$$

Pro každé pevné $a \in A$ je $(P_{Y|X}(a, b))_{b \in B}$ pravděpodobnostní rozdělení na B. Jeho entropii označujeme $\mathcal{H}(Y | X = a)$.

$$\mathcal{H}(Y | X = a) = \sum_{b \in B} \frac{P_{XY}(a, b)}{P_X(a)} \cdot \log \frac{P_X(a)}{P_{XY}(a, b)}$$

Podmíněná entropie $\mathcal{H}(Y | X = a)$ může být k entropii $\mathcal{H}(Y)$ v libovolném vztahu. Dozvíme-li se hodnotu náhodné proměnné X může se naše nejistota o Y jak zvýšit tak snížit.

Příklad 4 *Uvažujme pravděpodobnostní rozdělení*

$$P_{XY} = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/4 \end{bmatrix}, \quad P_{Y|X} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}, \quad P_{X|Y} = \begin{bmatrix} 2/3 & 0 \\ 1/3 & 1 \end{bmatrix}$$

Pak $P_X = [\frac{1}{2}, \frac{1}{2}]^T$, $P_Y = [\frac{3}{4}, \frac{1}{4}]$, $\mathcal{H}(X) = 1$, $\mathcal{H}(Y) = 0.811$, $\mathcal{H}(X,Y) = \frac{3}{2}$, $\mathcal{H}(X:Y) = 0.311$. Pro podmíněné entropie dostáváme

$$0 = \mathcal{H}(Y|X=0) < \mathcal{H}(Y) < \mathcal{H}(Y|X=1) = 1.$$

Podmíněná entropie $\mathcal{H}(Y|X) = \frac{1}{2}$ je již menší než $\mathcal{H}(Y)$. Podobně $\mathcal{H}(X|Y) = 0.689 < \mathcal{H}(X)$.

Definice 7 **Informační obsah** *a* **entropie** *náhodné veličiny X za* **podmínky** *Y je*

$$\begin{aligned}
\mathfrak{I}_{Y|X} &= -\sum_{a\in A}\sum_{b\in B} \log \mathbb{P}[Y=b|X=a] \cdot \chi_{X^{-1}(a)\cap Y^{-1}(b)} \\
\mathcal{H}(Y|X) &= \mathbb{E}(\mathfrak{I}_{Y|X}) = \sum_{a\in A} P_X(a)\cdot \mathcal{H}(Y|X=a) = \sum_{a\in A}\sum_{b\in B} P_{XY}(a,b)\cdot \log \frac{P_X(a)}{P_{XY}(a,b)}
\end{aligned}$$

Je tedy $\mathfrak{I}_{Y|X}(\omega) = -\log P_{Y|X}(X(\omega), Y(\omega))$. Podmíněná informace je průměrné množství informace nutné k určení Y, pokud již známe X.

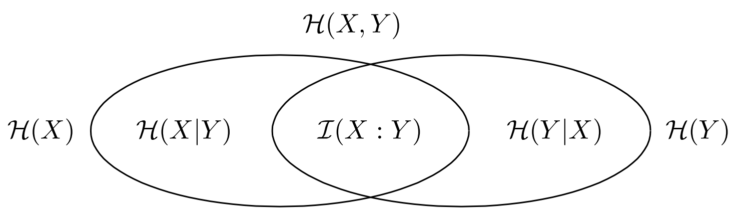
Tvrzení 14 *Nechť X, Y jsou náhodné veličiny. Pak*

(1) $\mathfrak{I}_{Y|X} = \mathfrak{I}_{X,Y} - \mathfrak{I}_X$

(2) $\mathcal{H}(Y|X) = \mathcal{H}(X,Y) - \mathcal{H}(X)$

(3) $\mathcal{H}(Y|X) = \mathcal{H}(Y) - \mathcal{I}(X:Y) \leq \mathcal{H}(Y)$

(4) $\mathcal{H}(X,Y) = \mathcal{H}(Y|X) + \mathcal{I}(X:Y) + \mathcal{H}(X|Y)$



Obrázek 9: Podmíněná a vzájemná entropie

Důkaz: S použitím Definice 6 dostáváme

$$\begin{aligned}
\mathfrak{I}_{Y|X}(\omega) &= -\log P_{Y|X}(X(\omega), Y(\omega)) = \log P_X(X(\omega)) - \log P_{X,Y}(X(\omega), Y(\omega)) \\
&= -\mathfrak{I}_X(\omega) + \mathfrak{I}_{X,Y}(\omega) \\
\mathcal{H}(Y|X) &= \sum_{a\in A}\sum_{b\in B} P_{XY}(a,b)\cdot \log P_X(a) - \sum_{a\in A}\sum_{b\in B} P_{XY}(a,b)\cdot \log P_{XY}(a,b) \\
&= -\mathcal{H}(X) + \mathcal{H}(X,Y) = \mathcal{H}(Y) - \mathcal{I}(X:Y) \leq \mathcal{H}(Y) \\
\mathcal{H}(X,Y) &= \mathcal{H}(Y) + \mathcal{H}(X|Y) = \mathcal{H}(Y|X) + \mathcal{I}(X:Y) + \mathcal{H}(X|Y) \ \square
\end{aligned}$$

Náhodné veličiny $X : \Omega \to A$, $Y : \Omega \to B$ a $Z : \Omega \to C$ jsou nezávislé, jestliže pro každé $a \in A$, $b \in B$, $c \in C$ platí $P_{XYZ}(a,b,c) = P_X(a)P_Y(b)P_Z(c)$. V tomto případě platí

$$\mathfrak{I}_{X,Y,Z} = \mathfrak{I}_X + \mathfrak{I}_Y + \mathfrak{I}_Z, \quad \mathcal{H}(X,Y,Z) = \mathcal{H}(X) + \mathcal{H}(Y) + \mathcal{H}(Z)$$

Příklad 5 *Nechť* $X, Y : \Omega \to B$ *jsou nezávislé náhodné veličiny s rozdělením* $(\frac{1}{2}, \frac{1}{2})$ *a* $Z = (X + Y) \mod 2$.

Pak X, Z jsou nezávislé, Y, Z jsou nezávislé, ale X, Y, Z nezávislé nejsou. Pro entropii platí

$$\mathcal{H}(X,Y,Z) = \mathcal{H}(X,Y) + \mathcal{H}(Z|X,Y) = \mathcal{H}(X,Y) = 2\mathcal{H}(X) = 2$$

Tvrzení 15 *Nechť* X, Y, Z *jsou náhodné veličiny. Pak*

(1) $\mathfrak{I}_{X,Y,Z} = \mathfrak{I}_X + \mathfrak{I}_{Y|X} + \mathfrak{I}_{Z|X,Y}$

(2) $\mathfrak{I}_{Y,Z|X} = \mathfrak{I}_{Y|X} + \mathfrak{I}_{Z|X,Y}$

(3) $\mathcal{H}(X,Y,Z) = \mathcal{H}(X) + \mathcal{H}(Y|X) + \mathcal{H}(Z|X,Y)$

(4) $\mathcal{H}(Y,Z|X) = \mathcal{H}(Y|X) + \mathcal{H}(Z|X,Y)$

(5) $\mathcal{H}(Z|X,Y) \leq \mathcal{H}(Z|Y)$

Důkaz: Informaci tří náhodných veličin můžeme počítat dvojím způsobem:

$$\mathfrak{I}_X + \mathfrak{I}_{Y,Z|X} = \mathfrak{I}_{X,Y,Z} = \mathfrak{I}_{X,Y} + \mathfrak{I}_{Z|X,Y} = \mathfrak{I}_X + \mathfrak{I}_{Y|X} + \mathfrak{I}_{Z|X,Y}$$

a odtud plyne (1) a (2). Přechodem ke střední hodnotě dostáváme (3) a (4). Pro důkaz (5) uvažujme pro pevné $b \in B$ náhodné veličiny X_b, Z_b s rozdělením

$$\begin{aligned}
\mathbb{P}[X_b = a, Z_b = c] &= \mathbb{P}[X = a, Z = c|Y = b] = P_{XYZ}(a,b,c)/P_Y(b) \\
\mathbb{P}[Z_b = c|X_b = a] &= \mathbb{P}[Z = c|X = a, Y = b] = P_{XYZ}(a,b,c)/P_{XY}(a,b)
\end{aligned}$$

Pak $\mathcal{H}(Z_b|X_b) \leq \mathcal{H}(Z_b)$ a tedy

$$\begin{aligned}
\mathcal{H}(Z|X,Y) &= \sum_{a,b,c} P_{XYZ}(a,b,c) \cdot \log \frac{P_{XY}(a,b)}{P_{XYZ}(a,b,c)} \\
&= \sum_{a,b,c} P_Y(b) \cdot \mathbb{P}[X_b = a, Z_b = c] \cdot \log \frac{1}{\mathbb{P}[Z_b = c|X_b = a]} \\
&= \sum_{b \in B} P_Y(b) \cdot \mathcal{H}(Z_b|X_b) \leq \sum_{b \in B} P_Y(b)\mathcal{H}(Z_b) = \mathcal{H}(Z|X) \ \square
\end{aligned}$$

Říkáme že náhodné veličiny X, Y, Z tvoří **markovský proces** $X \to Y \to Z$, pokud platí

$$\mathbb{P}[Z = c|X = a, Y = b] = \mathbb{P}[Z = c|Y = b]$$

Tvrzení 16 *Jestliže* X, Y, Z *tvoří markovský proces, pak* $\mathcal{H}(Z|X,Y) = \mathcal{H}(Z|Y)$.

Důkaz:

$$\begin{aligned}
\mathcal{H}(Z|X,Y) &= -\sum_{a \in A} \sum_{b \in B} P_{X,Y}(a,b) \sum_{c \in C} \mathbb{P}[Z = c|X = a, Y = b] \log \mathbb{P}[Z = c|X = a, Y = b] \\
&= -\sum_{a \in A} \sum_{b \in B} P_{X,Y}(a,b) \sum_{c \in C} \mathbb{P}[Z = c|Y = b] \log \mathbb{P}[Z = c|Y = b] \\
&= \mathcal{H}(Z|Y) \ \square
\end{aligned}$$

Tvrzení 17 *Pro posloupnost náhodných veličin* $X_0, \ldots X_{n-1}$ *platí*

$$\begin{aligned}
\mathcal{H}(X_{[0,n)}) &= \sum_{i=0}^{n-1} \mathcal{H}(X_i|X_{[0,i)}) \\
&= \mathcal{H}(X_0) + \mathcal{H}(X_1|X_0) + \mathcal{H}(X_2|X_{[0,1]}) + \cdots + \mathcal{H}(X_{n-1}|X_{[0,n-1)})
\end{aligned}$$

2 Náhodné procesy

Text psaný buď v přirozeném nebo programovacím jazyce modelujeme jako náhodný proces, tj. posloupnost náhodných veličin. Tyto náhodné veličiny nejsou nezávislé. Pravděpodobnoast výskytu písmene je značně ovlivněna kontextem v jakém se nachází. Tato závislost snižuje entropii textu a umožňuje jeho kompresi.

Definice 8 **Náhodný proces** *nad abecedou A je nekonečná posloupnost náhodných veličin $X = (X_n : \Omega \to A)_{n\in\mathbb{N}}$. Rozdělení procesu je zobrazení $\mathcal{P}_X : A^* \to [0,1]$*

$$\mathcal{P}_X(u) = \mathbb{P}[X_{[0,|u|)} = u_{[0,|u|)}] = \mathbb{P}[X_0 = u_0, \ldots, X_{|u|-1} = u_{|u|-1}], \;\; u \in A^*$$

Proces X je **stacionární***, pokud má stejné rozdělení jako posunutý proces $(X_i)_{i\geq 1}$.*

Pro zobrazení $\mathcal{P}_X$ platí Kolmogorovy podmínky kompatibility

$$\sum_{a\in A} \mathcal{P}_X(ua) = \sum_{a\in A} \mathbb{P}[X_{[0,|u|)} = u, X_{|u|} = a] = \mathbb{P}[X_{[0,|u|)} = u] = \mathcal{P}_X(u), \; u \in A^*$$

a $\mathcal{P}_X(\lambda) = \mathbb{P}(\Omega) = 1$. Je-li proces X stacionární, platí navíc

$$\sum_{a\in A} \mathcal{P}_X(au) = \sum_{a\in A} \mathbb{P}[X_0 = a, X_{[1,|u|]} = u] = \mathbb{P}[X_{[1,|u|]} = u] = \mathbb{P}[X_{[0,|u|)} = u] = \mathcal{P}_X(u).$$

Definice 9 *Symbolická míra je zobrazení $\mu : A^* \to [0,1]$, které splňuje* **podmínky kompatibility**

$$\mu(\lambda) = 1, \;\; \sum_{a\in A} \mu(ua) = \mu(u), \;\; u \in A^*$$

Stacionární symbolická míra *je zobrazení $\mu : A^* \to [0,1]$, které splňuje* **oboustranné podmínky kompatibility**

$$\mu(\lambda) = 1, \;\; \sum_{a\in A} \mu(au) = \sum_{a\in A} \mu(ua) = \mu(u), \;\; u \in A^*$$

Tvrzení 18 *Je-li X náhodný proces, je $\mathcal{P}_X$ symbolická míra. Proces X je stacionární právě když $\mathcal{P}_X$ je stacionární míra. Je-li $\mu : A^* \to [0,1]$, symbolická míra, existuje náhodný proces X takový že $\mathcal{P}_X = \mu$.*

Důkaz: Položme $\Omega = A^{\mathbb{N}}$, kde $A^{\mathbb{N}}$ je množina nekonečných slov $x = x_0x_1x_2\ldots$ v abecedě A. Nechť $\mathcal{B}$ je σ-algebra borelovských množin. Ta je definována jako nejmenší σ-algebra, která obsahuje všechny cylindry, tj. množiny tvaru

$$[u] = \{x \in A^{\mathbb{N}} : \; x_{[0,|u|)} = u\}, \;\; u \in A^*$$

Symbolickou míru μ lze definovat na cylindrech jako $\mu([u])\mu(u)$ a jednoznačně rozšířit na pravděpodobnostní míru, tj. na spočetně-aditivní zobrazení $\mu : \mathcal{B} \to [0,1]$. Náhodnou veličinu $X_i : A^{\mathbb{N}} \to A$ definujeme jako projekci $X_i(x) = x_i$. Pak $X = (X_i : A^{\mathbb{N}} \to A)_{i\geq 0}$ je náhodný proces s rozdělením $\mathcal{P}_X = \mu$. $\square$.

Je-li μ symbolická míra, je její restrikce na A^n rozdělení které značíme $\mu_{|A^n} \in \Delta(A^n)$. Je-li $P \in \Delta(A)$ rozdělení, označme $P^n \in \Delta(A^n)$ rozdělení

$$P^n(u) = P(u_0)\cdots P(u_{n-1}), \;\; u \in A^n$$

Pro $M \subseteq A^n$ značíme

$$P^n(M) = \sum_{u\in M} P^n(u)$$

Pro dané $P \in \Delta(A)$ položme

$$\mu(u) = P^{|u|}(u), \quad u \in A^*$$

Pak μ je stacionární míra, která se nazývá **bernoulliovská**. Bernoulliovské míry jsou právě rozložení procesů tvořených stejně rozdělenými navzájem nezávislými náhodnými veličinami.

Definice 10 *Dolní a horní entropie náhodného procesu je*

$$\underline{\mathcal{H}}(X) = \liminf_{n\to\infty} \mathcal{H}(X_{[0,n)})/n, \quad \overline{\mathcal{H}}(X) = \limsup_{n\to\infty} \mathcal{H}(X_{[0,n)})/n.$$

Pokud dolní a horní entropie splývá, říkáme že proces má entropii $\mathcal{H}(X) = \underline{\mathcal{H}}(X) = \overline{\mathcal{H}}(X)$.

Věta 19 *Stacionární proces X má entropii*

$$\mathcal{H}(X) = \lim_{n\to\infty} \mathcal{H}(X_n|X_{[0,n)}) = \lim_{n\to\infty} \frac{1}{n}\mathcal{H}(X_{[0,n)}).$$

Důkaz: Označme $a_0 = \mathcal{H}(X_0)$, $a_n = \mathcal{H}(X_n|X_{[0,n)})$. Podle Tvrzení 17

$$a_n \le \mathcal{H}(X_n|X_{[1,n)}) = \mathcal{H}(X_{n-1}|X_{[0,n-1)}) = a_{n-1}$$

takže a_n je nerostoucí posloupnost a její limita $a = \lim_{n\to\infty} a_n$ existuje. Podle Tvrzení 17 je

$$b_n = \frac{\mathcal{H}(X_{[0,n)})}{n} = \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{H}(X_i|X_0,\ldots,X_{i-1}) = \frac{1}{n}\sum_{i=0}^{n-1} a_i$$

Pro každé $\varepsilon > 0$ existuje m takové že pro všechna $n \ge m$ je $|a_n - a| < \varepsilon$ a tedy

$$|b_n - a| \le \frac{1}{n}\left(\sum_{i=0}^{m-1} |a_i - a| + \sum_{i=m}^{n-1} |a_i - a|\right) \le \frac{1}{n}\sum_{i=0}^{m-1} |a_i - a| + \varepsilon$$

takže $\limsup_{n\to\infty} |b_n - a| \le \varepsilon$. Protože to platí pro každé $\varepsilon > 0$, je $\lim_{n\to\infty} b_n = a$. $\square$

Je-li X stacionární proces s rozdělením $\mathcal{P}_X$, pak $\overline{\mathcal{P}}(u_0 \ldots u_{n-1}) = \mathcal{P}_X(u_{n-1} \ldots u_0)$ splňuje oboustranné podmínky kompatibility a tedy určuje stacionární proces. Říkáme že proces Y je inverzní k X, pokud mají navzájem inverzní rozdělení. Operace inverze reprezentuje obrácení směru času.

Tvrzení 20 *Je-li X stacionární, je*

$$\mathcal{H}(X_0|X_{(0,n]}) = \mathcal{H}(X_n|X_{[0,n)}), \quad \mathcal{H}(X_0|X_n) = \mathcal{H}(X_n|X_0)$$

Důkaz: $\mathcal{H}(X_0|X_{(0,n]}) = \mathcal{H}(X_{[0,n]}) - \mathcal{H}(X_{(0,n]}) = \mathcal{H}(X_{[0,n]}) - \mathcal{H}(X_{[0,n)}) = \mathcal{H}(X_n|X_{[0,n)})$. Podobně platí $\mathcal{H}(X_0|X_n) = \mathcal{H}(X_0, X_n) - \mathcal{H}(X_n) = \mathcal{H}(X_0, X_n) - \mathcal{H}(X_0) = \mathcal{H}(X_n|X_0)$. $\square$

Tvrzení 21 *Navzájem inverzní stacionární procesy mají stejnou entropii.*

Důkaz: Nechť Y je proces inverzní k X.

$$\mathcal{H}(Y) = \lim_{n\to\infty} \frac{1}{n}\mathcal{H}(Y_{[0,n)}) = \lim_{n\to\infty} \frac{1}{n}\mathcal{H}(X_{[0,n)}) = \mathcal{H}(X) \ \square$$

2.1 Markovské procesy

Definice 11 **Markovský proces** *řádu* $k \geq 0$ *je náhodný proces pro který platí*

$$\mathbb{P}[X_n = u_n | X_{[0,n)} = u_{[0,n)}] = \mathbb{P}[X_n = u_n | X_{[n-k,n)} = u_{[n-k,n)}], \;\; n > k$$

Pro $k = 0$ má markovská podmínka tvar

$$\mathbb{P}[X_n = u_n | X_{[0,n-1)} = u_{[0,n-1)}] = \mathbb{P}[X_n = u_n],$$

takže markovský proces řádu 0 je posloupnost nezávislých náhodných veličin. Tyto náhodné veličiny nemusí mít nutně stejné rozdělení, v tomto případě však proces není stacionární. Stacionární markovský proces řádu 0 je posloupnost stejně rozdělených nezávislých náhodných veličin, tedy bernoulliovský proces. Markovský proces řádu $k \geq 2$ lze ekvivalentně popsat jako markovský proces prvního řádu:

Tvrzení 22 *Je-li* $X = (X_n : \Omega \to A)_{n \geq 0}$ *markovský proces v abecedě* A *řádu* $k \geq 2$*, je* $Y = (X_{[kn,kn+k)})_{n \geq 0}$ *markovský proces řádu* 1 *v abecedě* A^k*.*

Důkaz: S použitím vzorce $\mathbb{P}[Y, Z|X] = \mathbb{P}[Y|X] \cdot \mathbb{P}[Z|X, Y]$ dostáváme

$$\begin{aligned}
&\mathbb{P}[Y_n = u_{[kn,kn+k)} | Y_{[0,n)} = u_{[0,kn)}] \\
&\quad = \mathbb{P}[X_{[kn,kn+k)} = u_{[kn,kn+k)} | X_{[0,kn)} = u_{[0,kn)}] \\
&\quad = \prod_{i=0}^{k-1} \mathbb{P}[X_{kn+i} = u_{kn+i} | X_{[0,kn+i)} = u_{[0,kn+i)}] \\
&\quad = \prod_{i=0}^{k-1} \mathbb{P}[X_{kn+i} = u_{kn+i} | X_{[kn+i-k,kn+i)} = u_{[kn+i-k,kn+i)}] \\
&\quad = \mathbb{P}[X_{[kn,kn+k)} = u_{[kn,kn+k)} | X_{[kn-k,kn)} = u_{[kn-k,kn)}] \\
&\quad = \mathbb{P}[Y_n = u_{[kn,kn+k)} | Y_{n-1} = u_{[kn-k,kn)}] \;\; \square
\end{aligned}$$

Stacionární markovský proces prvního řádu lze popsat maticí přechodových pravděpodobností.

Definice 12 **Stochastická matice** *nad abecedou* A *je čtvercová matice* $R = (R(a,b))_{a,b \in A}$ *nezáporných čísel, která pro každé* $a \in A$ *splňuje* $\sum_{b \in A} R(a,b) = 1$*.*

Pro stacionární markovský proces X prvního řádu definujme jeho **stacionární vektor** $P \in \Delta(A)$ a **přechodovou matici** R předpisem

$$P(a) = \mathcal{P}_X(a), \;\; R(a,b) = \mathbb{P}[X_1 = b | X_0 = a] = \frac{\mathcal{P}_X(ab)}{\mathcal{P}_X(a)}$$

Pak R je stochastická matice a platí $P \cdot R = P$, tj.

$$\sum_{a \in A} P(a) \cdot R(a,b) = P(b)$$

Naopak je-li $P \in \Delta(A)$ pravděpodobnostní rozdělení, a R stochastická matice splňující $P \cdot R = P$, existuje stacionární markovský proces prvního řádu se stacionárním rozdělením P a přechodovou maticí R. Rozložení procesu je dáno vzorcem

$$\mu(u) = P(u_0) \cdot R(u_0, u_1) \cdots R(u_{|u|-2}, u_{|u|-1}).$$

Podmíněné pravděpodobnosti $\mathbb{P}[X_n = b|X_0 = a] = R^n(a,b)$ jsou určeny mocninami přechodové matice. Je totiž

$$\begin{aligned}\mathbb{P}[X_2 = c|X_0 = a] &= \sum_{b\in A} \mathbb{P}[X_1 = b|X_0 = a] \cdot \mathbb{P}[X_2 = c|X_0 = a, X_1 = b] \\ &= \sum_{b\in B} R(a,b) \cdot R(b,c) = R^2(a,c)\end{aligned}$$

Věta 23 *Entropie stacionárního markovského procesu X prvního řádu se stacionárním rozdělením P a přechodovou maticí R je*

$$\mathcal{H}(X) = \mathcal{H}(X_1|X_0) = -\sum_{a\in A} P(a) \sum_{b\in A} R(a,b) \cdot \log R(a,b).$$

Důkaz: $\mathcal{H}(X_n|X_{[0,n)}) = \mathcal{H}(X_n|X_{n-1}) = \mathcal{H}(X_1|X_0)$. □

Tvrzení 24 *Je-li X stacionární markovský proces řádu 1 s primitivní přechodovou maticí R a stacionárním vektorem P, je proces k němu inversní také stacionární markovský proces řádu 1 se stejným stacionárním vektorem P a přechodovou maticí*

$$\overline{R}(b,a) = P(a) \cdot R(a,b)/P(b)$$

Důkaz:

$$\begin{aligned}\mathbb{P}[X_0 = u_0|X_{[1,n]} = u_{[1,n]}] &= \frac{P(u_0)R(u_0,u_1)\cdots R(u_{n-1},u_n)}{P(u_1)R(u_1,u_2)\cdots R(u_{n-1},u_n)} = \frac{P(u_0)R(u_0,u_1)}{P(u_1)} \\ &= \overline{R}(u_1,u_0) = \mathbb{P}[X_0 = u_0|X_1 = u_1]\end{aligned}$$

Tvrzení 25 *Je-li X stacionární markovský proces řádu 1, pak*

(1) $\mathcal{H}(X_{i+1}|X_{[0,i]}) = \mathcal{H}(X_{i+1}|X_i)$.

(2) $\mathcal{I}(X_0 : X_n) \le \mathcal{I}(X_0 : X_{n-1})$.

(3) $\mathcal{H}(X_n|X_{[0,n)\cup(n,2n]}) = \mathcal{H}(X_1|X_0,X_2) = 2\mathcal{H}(X_1|X_0) - \mathcal{H}(X_2|X_0)$

Důkaz: (1) plyne bezprostředně z definice. Pro podmíněné pravděpodobnosti snadno odvodíme

$$\mathbb{P}[X_n = u_n|X_0 = u_0, X_1 = u_1] = \mathbb{P}[X_n = u_n|X_1 = u_1]$$

Takže $\mathcal{H}(X_n|X_0,X_1) = \mathcal{H}(X_n|X_1)$ a tedy

$$\begin{aligned}\mathcal{I}(X_0 : X_{n-1}) &= \mathcal{I}(X_1 : X_n) = \mathcal{H}(X_n) - \mathcal{H}(X_n|X_1) = \mathcal{H}(X_n) - \mathcal{H}(X_n|X_0,X_1) \\ &\ge \mathcal{H}(X_n) - \mathcal{H}(X_n|X_0) = \mathcal{I}(X_0 : X_n). \\ \mathcal{H}(X_{[0,n)\cup(n,2n]}) &= \mathcal{H}(X_0) + \mathcal{H}(X_1|X_0) + \ldots + \mathcal{H}(X_{n-1}|X_{n-2}) + \\ &\quad \mathcal{H}(X_{n+1}|X_{n-1}) + \mathcal{H}(X_{n+2}|X_{n+1}) + \ldots + \mathcal{H}(X_{2n}|X_{2n-1}) \\ &= \mathcal{H}(X_0) + 2(n-1)\mathcal{H}(X_1|X_0) + \mathcal{H}(X_2|X_0) \\ \mathcal{H}(X_n|X_{[0,n)\cup(n,2n]}) &= \mathcal{H}(X_{[0,2n]}) - \mathcal{H}(X_{[0,n)\cup(n,2n]}) = 2\mathcal{H}(X_1|X_0) - \mathcal{H}(X_2|X_0)\end{aligned}$$ □

Bod (3) Tvrzení 25 se vztahuje k situaci, kdy neznáme jedno písmeno textu a snažíme se ho určit z celého kontextu, tj. z písmen předcházejících i následujících.

Tvrzení 26 *Pro každou stochastickou matici R existuje stacionární vektor P, pro který $P \cdot R = P$.*

Důkaz: Prostor pravděpodobnostních vektorů $\Delta(A) \subset [0,1]^A$ s eukleidovskou metrikou je kompaktní. Zvolme libovolné $Q \in \Delta_A$ a položme $Q_n = (Q + Q \cdot R + \cdots + Q \cdot R^{n-1})/n$. Pak $Q_n \in \Delta(A)$ a existuje vybraná konvergentní posloupnost $\lim_{i\to\infty} Q_{n_i} = P$. Platí

$$P \cdot R - P = \lim_{i\to\infty}(Q_{n_i} \cdot R - Q_{n_i}) = \lim_{i\to\infty} \frac{1}{n_i}(Q \cdot R^{n_i} - Q) = 0. \ \square$$

Příklad 6 *Nechť* $0 < p, q < 1$ *a položme*

$$R = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \quad P = \left[\frac{q}{p+q}, \frac{p}{p+q}\right].$$

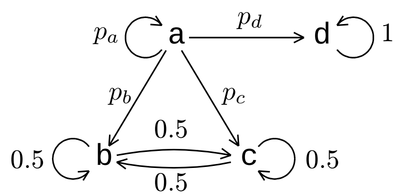
Pak P je jediné stacionární rozdělení R. Entropie příslušného stacionárního procesu je

$$\mathcal{H}(X) = \frac{q \cdot \mathcal{H}(p, 1-p) + p \cdot \mathcal{H}(q, 1-q)}{p+q}.$$

Stochastické matice zadáváme grafem jehož vrcholy jsou písmena abecedy a jehož hrany jsou označeny pravděpodobnostmi.

Příklad 7 *Nechť* $p = (p_a, p_b, p_c, p_d)$ *je kladné pravděpodobnostní rozdělení. Stochastická matice* R *na obrázku 10 má stacionární rozdělení tvaru* $P = [0, \frac{q}{2}, \frac{q}{2}, 1-q]$, *kde* $0 \le q \le 1$.

$$R = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & p_a & p_b & p_c & p_d \\ b & 0 & 1/2 & 1/2 & 0 \\ c & 0 & 1/2 & 1/2 & 0 \\ d & 0 & 0 & 0 & 1 \end{array}$$



Obrázek 10: Markovský proces s přechodným stavem

V tomto procesu je stav a **přechodný**: Je-li $X_0 = a$, proces po konečném počtu kroků přechází buď do stavu d s pravděpodobností $p_d/(1-p_a)$ a zde již zůstává, nebo do množiny stavů $\{b, c\}$ s pravděpodobností $(p_b + p_c)/(1-p_a)$. Je totiž

$$\begin{aligned} \mathbb{P}[\exists n, X_n \in \{b,c\} | X_0 = a] &= \sum_{n=1}^{\infty} p_a^{n-1}(p_b + p_c) = (p_b + p_c)/(1-p_a) \\ \mathbb{P}[\exists n, X_n = d | X_0 = a] &= \sum_{n=1}^{\infty} p_a^{n-1} p_d = p_d/(1-p_a) \end{aligned}$$

Rovnice $P \cdot R = P$ pro stacionární rozdělení dává $P(a) = 0$, $P(b) = P(c)$ a obecný tvar stacionárního rozdělení je $P = [0, \frac{q}{2}, \frac{q}{2}, 1-q]$. Entropie příslušného procesu závisí na parametru q:

$$\mathcal{H}(X) = q \cdot \mathcal{H}(0, \tfrac{1}{2}, \tfrac{1}{2}, 0) + (1-q) \cdot \mathcal{H}(0,0,0,1) = q.$$

Podmínku jednoznačnosti stacionárního rozdělení dává Perron-Frobeniova teorie nezáporných matic. Všimněme si že stacionární rozdělení P které splňuje $P \cdot R = P$ je levý vlastní vektor matice A, který přísluší její vlastní hodnotě 1. Každá stochastická matice R má vlastní hodnotu 1 protože má jednotkový pravý vlastní vektor $e = [1, \ldots, 1]$, tj. $R \cdot e^T = e^T$, kde e^T je sloupcový vektor transponovaný k řádkovému vektoru e.

2.2 Nezáporné matice

Říkáme že matice $M = (M_{ab})_{a,b\in A}$ je nezáporná ($M \geq 0$), jestliže $M_{ab} \geq 0$ pro všechna $a, b \in A$, a že je kladná ($M > 0$), pokud $M_{ab} > 0$ pro všechna $a, b \in A$. Podobně mluvíme o nezáporných a kladných vektorech $u = (u_a)_{a\in A}$. Na vektor se díváme jako na řádkový vektor tj. na matici typu $1 \times A$. Příslušný sloupcový vektor, tj. matici typu $A \times 1$ značíme u^T. Skalární součin vektorů $u, v \in \mathbb{R}^A$ je $uv^T = \sum_{a\in A} u_a v_a$. Naopak $u^T v$ je matice typu $A \times A$ a $(u^T v)_{ab} = u_a v_b$.

Definice 13 *Nechť $(M_{ab})_{a,b\in A}$ je nezáporná čtvercová matice.*

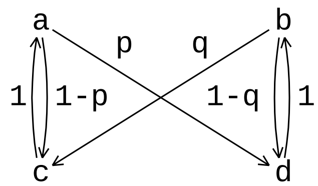
(1) M je ireducibilní, pokud pro každé $a, b \in A$ existuje $n > 0$ pro které $(M^n)_{ab} > 0$.

(2) M je primitivní, pokud existuje $n > 0$ takové že $M^n > 0$

Příklad 8 *Nechť $0 < p, q < 1$. Stochastická matice R na obrázku 11 je ireducibilní ale není primitivní. Má jediné stacionární rozdělení*

$$P = \left[\frac{q}{2(p+q)}, \frac{p}{2(p+q)}, \frac{q}{2(p+q)}, \frac{p}{2(p+q)}\right].$$

$$R = \begin{array}{c|cccc} & a & b & c & d \\ \hline a & 0 & 0 & 1-p & p \\ b & 0 & 0 & q & 1-q \\ c & 1 & 0 & 0 & 0 \\ d & 0 & 1 & 0 & 0 \end{array}$$



Obrázek 11: Ireducibilní stochastická matice

Komplexní číslo ϱ je vlastní hodnota $M = (M_{ab})_{a,b\in A}$, pokud existuje nenulový komplexní (levý) vlastní vektor $v = (v_a)_{a\in A}$ takový že $vM = \varrho v$ tj. $\sum_{a\in A} v_a M_{ab} = \varrho v_b$ pro každé $b \in A$. Je-li ϱ vlastní hodnota M a v levý vlastní vektor pak $v(M - \varrho I) = 0$, kde I je jednotková matice, takže ϱ je řešení algebraické rovnice $\det(M - \varrho I) = 0$ stupně $|A|$. Existuje nejvýše $|A|$ vlastních hodnot M. Vlastní hodnota ϱ má také pravý vlastní vektor u takový že $Mu^T = \varrho u^T$. Pro vektor $x \in \mathbb{R}^A$ nechť $|x| \in \mathbb{R}^A$ je vektor jeho absolutních hodnot, $|x|_a = |x_a|$ a $||x|| = \sqrt{\sum_{a\in A} x_a^2}$ jeho norma

Věta 27 (Perron-Frobeniova věta) *Nechť M je primitivní matice. Existuje $\varrho = \varrho(M) > 0$ které nazýváme spektrální poloměr M, a kladné vektory u, v pro které platí*

(1) $vM = \varrho v$ (ϱ je vlastní hodnota M a v je její levý vlastní vektor).

(2) Každý levý vlastní vektor ϱ je (komplexní) násobek v.

(3) Je-li μ vlastní hodnota M a $\mu \neq \varrho$, pak $|\mu| < \varrho$.

(4) Je-li u pravý vlastní vektor ϱ a $v \cdot u^T = \sum_{a\in A} u_a v_a = 1$, pak

$$\lim_{n\to\infty} \frac{M^n}{\varrho^n} = u^T \cdot v.$$

Důkaz: (1) Položme $\varrho = \sup\{a > 0 :\ \exists x \in \mathbb{R}^A, x \geq 0,\ ax \leq xM\}$. Pro $x = (1, \ldots, 1)$ platí $xM > 0$, takže $\varrho > 0$. Existuje posloupnost $(a_n)_{n\geq 0}$, která konverguje k ϱ a nezáporné vektory $x^{(n)} \in \mathbb{R}^A$ pro které $a_n x^{(n)} \leq x_n M$ a $||x_{(n)}|| = 1$. V kompaktním prostoru $\{x \in \mathbb{R}^A :\ x \geq 0, ||x|| = 1\}$ existuje konvergentní podposloupnost $\lim_{i\to\infty} x^{(n_i)} = v$ a $\varrho v \leq vM$. Předpokládejme že $w = vM - \varrho v \geq 0$ je nenulové. Existuje $p > 0$ pro které $M^p > 0$ a

$$(vM)M^p - (\varrho v)M^p = wM^p > 0,$$

takže existuje $\varrho' > \varrho$ pro které $\varrho'(vM^p) \leq (vM^p)M$ a to je spor. Je tedy $\varrho v = vM$ a v je levý vlastní vektor. Protože $\varrho^p v = vM^p > 0$, v je kladný.

(2) Předpokládejme že $wM = \varrho w$ a w je nezávislý s v. Můžeme předpokládat že w je reálný. Existuje lineární kombinace $z = cv + dw$ která je nezáporná ale ne kladná. Pak $\varrho^p z = zM^p > 0$, takže $z > 0$ a to je spor.

(3) Nechť μ je vlastní hodnota M a w její levý vlastní vektor tj. $wM = \mu w$. Jak μ tak w mohou být komplexní. Dostáváme

$$|\mu| \cdot |w_b| = \left| \sum_{a\in A} w_a M_{ab} \right| \leq \sum_{a\in A} |w_a| M_{ab}.$$

Z definice ϱ plyne $|\mu| \leq \varrho$. Předpokládejme že $|\mu| = \varrho$. Pak $\varrho|w| \leq |w|M$ a stejně jako v důkazu (1) dostáváme $\varrho|w| = |w|M$, takže $|w|$ je vlastní vektor a tedy násobek v. Je tedy

$$\sum_{a\in A} |w_a|(M^p)_{ab} = \varrho^p |w_b| = \left| \sum_{a\in A} w_a (M^p)_{ab} \right|.$$

Indukcí podle velikosti A lze dokázat že to je možné pouze pokud w lze získat z v násobením komplexní jednotkou, tj. $w_a = |v_a| \cdot e^{i\alpha}$. Je tedy $\varrho v = vM = \mu v$ takže $\mu = \varrho$ a to je spor.

(4) Pro nezáporný vektor x položme

$$r_0(x) = \min\left\{ \frac{x_a}{v_a} : a \in A \right\}, \quad r_1(x) = \max\left\{ \frac{x_a}{v_a} : a \in A \right\}.$$

Pro každé $a \in A$ dostáváme

$$\frac{(xM)_b}{\varrho v_b} = \sum_{a\in A} \frac{x_a}{v_a} \cdot \frac{v_a M_{ab}}{\varrho v_b} \geq r_0(x) \cdot \sum_{a\in A} \frac{v_a M_{ab}}{\varrho v_b} = r_0(x).$$

Podobně pro $r_1(x)$, takže

$$r_0(x) \leq r_0\left(\frac{xM}{\varrho}\right) \leq r_1\left(\frac{xM}{\varrho}\right) \leq r_1(x).$$

Existují limity monotóních posloupností

$$s_0(x) = \lim_{n\to\infty} r_0\left(\frac{xM^n}{\varrho^n}\right), \quad s_1(x) = \lim_{n\to\infty} r_1\left(\frac{xM^n}{\varrho^n}\right),$$

a $s_0(x) \leq s_1(x)$. Ukážeme $s_0(x) = s_1(x)$. Existuje p, pro které $M^p > 0$. Položme

$$m = \min\left\{ \frac{v_a (M^p)_{ab}}{v_b} : a, b \in A \right\} > 0$$

Nechť x je nezáporný vektor a $c \in A$ je takové, že $\frac{x_c}{v_c} = r_1(x)$. Pak

$$\begin{aligned}
\frac{(xM^p)_b}{\varrho^p v_b} &= \sum_{a\neq c} \frac{x_a}{v_a} \cdot \frac{v_a (M^p)_{ab}}{\varrho^p v_b} + \frac{x_c}{v_c} \cdot \frac{v_c (M^p)_{cb}}{\varrho^p v_b} \\
&\geq \sum_{a\neq c} \frac{v_a (M^p)_{ab}}{\varrho^p v_b} \cdot r_0(x) + \frac{v_c (M^p)_{ab}}{\varrho^p v_b} \cdot (r_0(x) + r_1(x) - r_0(x)) \\
&\geq r_0(x) + m(r_1(x) - r_0(x))
\end{aligned}$$

Je tedy

$$r_0\left(\frac{xM^p}{\varrho^p}\right) \quad \geq \quad (1-m)r_0(x)+mr_1(x)$$

Podobně dostáváme

$$\begin{aligned} r_1\left(\frac{xM^p}{\varrho^p}\right) \quad &\leq \quad (1-m)r_1(x)+mr_0(x) \\ r_1\left(\frac{xM^p}{\varrho^p}\right) - r_0\left(\frac{xM^p}{\varrho^p}\right) \quad &\leq \quad (1-2m)(r_1(x)-r_0(x)) \end{aligned}$$

Existují tedy limity

$$\lim_{n\to\infty}\frac{(xM^n)_b}{\varrho^n v_b} = s(x), \text{ neboli } \lim_{n\to\infty}\frac{xM^n}{\varrho^n} = vs(x).$$

Protože transponovaná matice M^T je také primitivní, má stejný spektrální poloměr ϱ, a kladný levý vlastní vektor u, který je pravým vlastním vektorem M. Je tedy $Mu^T = \varrho M$ a u můžeme normovat tak aby $v \cdot u^T = 1$. Pro $a \in A$, nechť e_a je jednotkový vektor tj. $(e_a)_a = 1$, $(e_a)_b = 0$ pro $b \neq a$. Pro $a, b \in A$ dostáváme

$$\begin{aligned} \lim_{n\to\infty}\frac{(M^n)_{ab}}{\varrho^n} \quad &= \quad \lim_{n\to\infty}\frac{(e_aM^n)_b}{\varrho^n} = s(e_a)v_b \\ \lim_{n\to\infty}\frac{(M^n)_{ab}}{\varrho^n} \quad &= \quad \lim_{n\to\infty}\frac{(M^ne_b)_a}{\varrho^n} = s'(e_b)u_a \end{aligned}$$

Existuje tedy $l > 0$ takové že pro všechna $a, b \in A$ platí $s(e_a)/u_a = s'(e_b)/v_b = l$. Protože stopa $\mathrm{tr}(M) = \sum_{a\in A} M_{aa}$ matice je součet jejích vlastních hodnot, dostáváme

$$1 = \lim_{n\to\infty}\sum_{a\in A}\frac{(M^n)_{aa}}{\varrho^n} = \sum_{a\in A} s(e_a)v_a = l\sum_{a\in A} v_au_a.$$

Je tedy $l = 1$ a z toho plyne výsledek. $\square$

Pro ireducibilní matice platí slabší verze Perron-Frobeniovy věty.

Věta 28 *Nechť M je ireducibilní matice. Existuje spektrální poloměr $\varrho = \varrho(M) > 0$ a kladný vektor v takový že*

(1) $vM = \varrho v$ (ϱ je vlastní hodnota M a v je její vlastní levý vlastní vektor).

(2) Každý levý vlastní vektor ϱ je násobek v.

(3) Je-li μ vlastní hodnota M, pak $|\mu| \leq \varrho$.

(4) Existují $0 < c_0 \leq 1 \leq c_1$ taková že pro každé $n > 0$ a $a \in A$,

$$c_0\varrho^n \leq \sum_{b\in A}(M^n)_{ab} \leq c_1\varrho^n.$$

Důkaz: (1) Pro každé $\varepsilon > 0$ je $M + \varepsilon I$ primitivní matice. Pokud $(M^n)_{ab} > 0$, pak $(M + \varepsilon I)^{n+m}_{ab} > 0$ pro každé $m \geq 0$, protože $M + \varepsilon I$ má kladnou diagonálu. Platí

$$vMu = \varrho v \iff v(M+\varepsilon I) = (\varrho+\varepsilon)v.$$

Pro $w = (1, \ldots, 1)$ je $w(M + \varepsilon I) > \varepsilon w$, takže spektrální poloměr ϱ_ε matice $M + \varepsilon I$ je větší než ε. Dále $\varrho = \varrho_\varepsilon - \varepsilon$ je kladná vlastní hodnota M a v je její kladný levý vlastní vektor.

(2) Je-li w levý vlastní vektor ϱ pro M pak je také vlastní vektor $\varrho+\varepsilon$ matice $M+\varepsilon I$, takže je násobkem v.
(3) Je-li μ vlastní hodnota M, pak $\mu+\varepsilon$ je vlastní hodnota $M+\varepsilon I$, takže $|\mu+\varepsilon|<\varrho+\varepsilon$. Protože to platí pro každé $\varepsilon>0$, je $|\mu|\leq\varrho$.
(4) Položme

$$d_0=\min\{u_a:\ a\in A\},\ \ d_1=\max\{u_a:\ a\in A\}.$$

Pro každé $a\in A$,

$$\frac{d_0}{d_1}\varrho^n\leq\frac{\varrho^n u_a}{d_1}=\sum_{b\in A}\frac{(M^n)_{ab}u_b}{d_1}\leq\sum_{b\in A}(M^n)_{ab}\leq\sum_{b\in A}\frac{(M^n)_{ab}u_b}{d_0}=\frac{\varrho^n u_a}{d_0}\leq\frac{d_1}{d_0}\varrho^n.\ \square$$

Věta 29 *Nechť M je ireducibilní matice která není primitivní. Pak existuje perioda $p>1$ a disjunktní rozklad abecedy $A=A_0\cup\cdots\cup A_{p-1}$ takový že*

$$a\in A_i\ \&\ M_{ab}>0\implies b\in A_{(i+1)\bmod p}$$

Pro každé $k<p$ je matice M^p omezená na A_k primitivní.

Důkaz: Pro $a\in A$ položme $P_a=\{n>0:\ (M^n)_{aa}>0\}$ a $p_a=\gcd P_a$ (největší společný dělitel). Množina P_a je uzavřená na sčítání a existuje $k>0$ takové že pro každé $n>k$ je $np_a\in P_a$. Protože M je ireducibilní, pro každé $a,b\in A$ existuje k,m taková že pro všechna $n>k$ je $m+np_a\in P_b$. Z toho plyne $p_a=p_b$. Periodu M definujeme tedy jako $p=p_a$ nejmenší společný dělitel libovolné P_a. Na A definujeme relaci

$$a\approx b\iff\exists k>0,(M^{kp})_{ab}>0$$

Pak $\approx$ je ekvivalence. Zvolme jako A_0 jednu třídu ekvivalence a pro $i<p$ položme

$$A_i=\{b\in A:\ \exists a\in A_0,\exists k>0,(M^{kp+i})_{ab}>0\}$$

Pak A_i je také třída ekvivalence a $A=A_0\cup\cdots\cup A_{p-1}$ je disjunktní rozklad A. Z definice plyne že každá matice $(M^p)_{a,b\in A_i}$ je primitivní. $\square$

Věta 30 *Nechť M je ireducibilní nezáporná matice, ϱ její spektrální poloměr a v,u příslušný levý a pravý vlastní vektor. Pak*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k<n}\frac{M^k}{\varrho^k}=u^T\cdot v$$

Důkaz: Nechť p je perioda M, ϱ je její spektrální poloměr a v,u její levý a pravý vlastní vektor. Protože $vM^p=\varrho^p v$, $M^p u^T=\varrho^p u^T$, je ϱ^p vlastní hodnota M^p a v,u její levý a vlastní vektor (matice M^p není ireducibilní). Nechť C je třída ekvivalence $\approx$. Pak M^p omezená na C je primitivní a vektory v,u omezené na C jsou její levý a pravý vlastní vektor. Z toho plyne $\lim_{q\to\infty}M^{qp}\cdot\varrho^{-qp}=u^T\cdot v$. Pro dané n pišme $n=qp+r$, kde $r<p$. Pak

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k<n}\frac{M^k}{\varrho^k} &= \lim_{n\to\infty}\frac{1}{n}\left(\sum_{i<p}\sum_{j<q}\frac{M^{i+pj}}{\varrho^{i+pj}}+\sum_{m<r}\frac{M^{m+pq}}{\varrho^{m+pq}}\right)\\ &= \sum_{i<p}\frac{M^i}{\varrho^i}\lim_{q\to\infty}\frac{1}{qp}\sum_{j<q}\frac{M^{pj}}{\varrho^{pj}}=\frac{1}{p}\sum_{i<p}\frac{M^i}{\varrho^i}u^T\cdot v\\ &= \frac{1}{p}\sum_{i<p}u^T\cdot v=u^T\cdot v\ \square\end{aligned}$$

Nechť $M = (M_{ab})_{a,b\in A}$ je nezáporná matice. Relace dosažitelnosti $\mathcal{P} \subseteq A \times A$ je

$$(a,b) \in \mathcal{P} \quad \Leftrightarrow \quad \exists n > 0, (M^n)_{ab} > 0$$

Relace $\mathcal{P}$ je tranzitivní. Relace $\mathcal{P}\cap\mathcal{P}^{-1}$ je tranzitivní a symetrická, není však nutně reflexivní. Označme $|\mathcal{P}| = \{a \in A : \ (a,a) \in \mathcal{P}\}$. Prvky $|\mathcal{P}|$ se nazývají rekurentní a prvky $A \setminus |\mathcal{P}|$ se nazývají tranzientní. Relace $\mathcal{P} \cap \mathcal{P}^{-1}$ omezená na $|\mathcal{P}|$ je relace ekvivalence. Třídy této ekvivalence se nazývají souvislé komponenty. Je-li $C \subseteq A$ souvislá komponenta, je matice $(M_{ab})_{a,b\in C}$ ireducibilní.

Definice 14 *Spektrální poloměr nezáporné matice $M = (M_{ab})_{a,b\in A}$ je maximum spektrálních poloměrů matic jejích souvislých komponent*

$$\varrho(M) = \max\{\varrho((M_{ab})_{a,b\in C}) : \ C \subseteq A \ \textit{ je souvislá komponenta}\}.$$

Pokud M žádnou souvislou komponentu nemá, klademe $\varrho(M) = 0$.

Tvrzení 31 *Vlastní hodnoty nezáporné matice sestávají z vlastních hodnot jejích souvislých komponent a (případně) z nuly, jejíž násobnost je počet tranzientních stavů.*

Důkaz: Nechť $M = (M_{ab})_{a,b\in A}$ je nezáporná matice. Abecedu A lze disjunktně rozložit na $A = A_0 \cup \cdots \cup A_{m-1}$, kde každé A_i je buď souvislá komponenta nebo jednobodová množina sestávající z tranzientního stavu. Definujme orientovaný graf, jehož vrcholy jsou A_i a

$$A_i \to A_j \iff \exists a \in A_i, \exists b \in A_j, M_{ab} > 0$$

Tento graf nemá cykly a množiny A_i lze uspořádat tak že pokud $A_i \to A_j$, pak $i < j$. Nechť $\pi : A \to A$ je permutace taková že $\prod_{a\in A}(M - \varrho I)_{a,\pi(a)} \neq 0$. Pak $\pi(A_{m-1}) \subseteq A_{m-1}$ a tedy $\pi(A_{m-1}) = A_{m-1}$. Protože $\pi(A_{m-2}) \subseteq A_{m-2} \cup A_{m-1}$, je $\pi(A_{m-2}) = A_{m-2}$ a podobně $\pi(A_i) = A_i$ pro každé $i < m$. Z toho plyne

$$\det(M - \varrho I) = \prod_{i<m} \det(M(i) - \varrho I(i))$$

kde $M(i), I(i)$ jsou matice M, I omezené na A_i. Je-li a tranzientní stav, je $M_{aa} = 0$ a z toho již plyne tvrzení. $\square$

Má-li matice A vlastní hodnotu ϱ násobnosti $m > 1$, označme $L(\varrho)$ čtvercovou matici typu $m \times m$

$$L(\varrho) = \begin{bmatrix} \varrho & 1 & 0 & \dots & 0 \\ 0 & \varrho & 1 & \dots & 0 \\ 0 & 0 & \varrho & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \varrho \end{bmatrix}, \text{ tj. } L(\varrho)_{ij} = \begin{cases} \varrho \text{ pro } i = j \\ 1 \text{ pro } j = i+1 \\ 0 \text{ pro } j < i \text{ nebo } j > i+1 \end{cases}$$

Nechť A má vlastní hodnoty ϱ_i s násobnostmi m_i, kde $1 \leq i \leq r$ a $m_1 + \cdots + m_r = n$. V tomto případě lze matici A rozložit na kanonický Jordanův tvar $A = V\Lambda U$, kde U, V jsou navzájem inverzní matice, a Λ je sestavena z matic $L(\varrho_i)$ na diagonále

$$\Lambda_{m_1+\ldots+m_{k-1}+i, m_1+\ldots+m_{k-1}+j} = L(\varrho_k)_{ij}, \ \ \Lambda_{ij} = 0 \text{ jinak}$$

Indukcí lze dokázat, že mocniny matice $L(\varrho)$ jsou

$$L(\varrho)^k_{ij} = \begin{cases} \binom{k}{j-i}\varrho^{k+i-j} & \text{pro } i \leq j \leq i+k \\ 0 & \text{pro } j < i \text{ nebo } j > i+k \end{cases}$$

Například pro $m = 3$ tyto mocniny jsou

$$\begin{bmatrix} \varrho & 1 & 0 \\ 0 & \varrho & 1 \\ 0 & 0 & \varrho \end{bmatrix}, \begin{bmatrix} \varrho^2 & 2\varrho & 1 \\ 0 & \varrho^2 & 2\varrho \\ 0 & 0 & \varrho^2 \end{bmatrix}, \begin{bmatrix} \varrho^3 & 3\varrho^2 & 3\varrho \\ 0 & \varrho^3 & 3\varrho^2 \\ 0 & 0 & \varrho^3 \end{bmatrix}, \ldots$$

Věta 32 *Nechť M je nezáporná matice se spektrálním poloměrem $\varrho > 0$ a nechť m je počet souvislých komponent, jejichž spektrální poloměr je ϱ. Pak existuje konstanta $c > 0$ taková že pro každé $a, b \in A$ a $n > 0$ platí $M^n_{ab} \leq c \cdot n^{m-1} \cdot \varrho^n$.*

Důkaz: Spektrální poloměr ϱ má násobnost m. Každý člen matice $L(\varrho)^n$ je nejvýše $n^{m-1}\varrho^n$. Pro každou jinou vlastní hodnotu μ je $|\mu| < \varrho$ a každý člen matice $L(\mu)^n$ je v absolutní hodnotě asymptoticky menší než ϱ^n. □

Tvrzení 33 *Každá primitivní stochastická matice R má spektrální poloměr 1 a jediné stacionární rozdělení P. Pro každé $a, b \in A$ platí $\lim_{n\to\infty} R^n(a, b) = P(b)$.*

Důkaz: Nechť ϱ je spektrální poloměr R. Protože 1 je vlastní hodnota příslušná jednotkovému pravému vlastnímu vektoru $u = (1, \ldots, 1)$, je $\varrho \geq 1$. Předpokládejme, že $\varrho > 1$ a v je její kladný levý vlastní vektor. Pak $v \cdot u^T > 0$ a $\varrho v \cdot u^T = v \cdot R \cdot u^T = v \cdot u^T$ a to je spor, takže $\varrho = 1$. Existuje tedy jediný levý kladný vlastní vektor $v = P$, pro který $P \cdot e^T = 1$ a

$$\lim_{n\to\infty} R^n(a, b) = u(a)v(b) = P(b). \quad \square$$

Věta 34 *Ireducibilní stochastická matice R má jediné stacionární rozdělení P. Pro každé $a, b \in A$ platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} R^i(a, b) = P(b).$$

Důkaz plyne z Věty 30.

Nechť $R = (R_{ab})_{a,b\in A}$ je nyní libovolná stochastická matice a $\mathcal{P}$ je její relace dosažitelnosti, tj. $(a, b) \in \mathcal{P}$ právě když $R^n_{ab} > 0$ pro nějaké $n > 0$. Říkáme že třída ekvivalence $C \subseteq |\mathcal{P}|$ je terminální, jestliže z ní (v grafu souvislých komponent) nevede žádá hrana. C je tedy terminální množina právě když platí

$$(\forall a, b \in C, \exists n > 0, R^n_{ab} > 0) \quad \& \quad (a \in C \;\&\; R_{ab} > 0 \implies b \in C)$$

Je-li C terminální komponenta, pak R omezená na C je stochastická matice. Označme $\mathcal{C}$ množinu terminálních komponent. Nechť X_i je (nestacionární) markovský proces prvního řádu s maticí přechodových pravděpodobností R s rovnoměrným rozdělením na A v čase 0. Pro $a \in A$, $C \in \mathcal{C}$ položme

$$\begin{aligned} P(C, a) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \mathbb{P}[X_k = a | X_0 \in C] \\ Q(a, C) &= \mathbb{P}[\exists n > 0, X_n \in C | X_0 = a] \end{aligned}$$

Matici P typu $\mathcal{C} \times A$ nazýváme stacionární a matici Q typu $A \times \mathcal{C}$ nazýváme absorpční. Platí $P(C, a) = 0$ pro $a \notin C$ a $(P(C, a))_{a\in C}$ je stacionární rozložení R na C. Pro $a \in C' \neq C$ platí $Q(a, C) = 0$.

Věta 35 *Nechť R je libovolná stochastická matice a P, Q její stacionární a absorpční matice. Pak P, Q jsou obě stochastické matice, $P \cdot R = P$, $R \cdot Q = Q$ a*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} R^k = Q \cdot P$$

Důkaz: Protože $(P(C,a))_{a\in C}$ je stacionární rozložení na C, je $\sum_{a\in A} P(C,a) = 1$, takže P je stochastická matice a platí $P \cdot R = P$. Nechť $A = A_0 \cup \cdots \cup A_{m-1}$ je rozklad abecedy A takový že každé A_i je buď souvislá komponenta nabo jednobodová množina sestávající z tranzientního stavu. Předpokládejme dále že pro $A_i \to A_j$ je $i < j$ a že terminální komponenty jsou na konci seznamu, tj $A_q, \cdots, A_{m-1}$, kde $q < m$. Je-li A_i neterminální souvislá komponenta, pak její spektrální poloměr je ostře menší než 1. Položme $B = A_0 \cup \cdots \cup A_{q-1}$. Pak R omezená na B má spektrální poloměr $\varrho < 1$. Zvolme $\varrho < \delta < 1$. Existuje $c > 0$ takové že pro každé $a \in B$ a každé $n > 0$ platí

$$\mathbb{P}[\forall m, X_m \in B | X_0 = a] \quad \leq \quad \sum_{m\geq n} \mathbb{P}[X_n \in B | X_0 = a] \leq \sum_{m \geq n} C\delta^m = \frac{C\delta^m}{1-\delta}$$

To znamená že $\sum_{C\in\mathcal{C}} Q(a,C) = 1$, takže Q je stochastická. Z definice Q plyne $R \cdot Q = Q$. □

Uvažujme nyní pro danou primitivní stochastickou matici R nad A a libovolné počáteční rozdělení $Q \in \Delta(A)$, nestacionární markovský proces Y s rozdělením

$$\mathcal{P}_Y(u) = Q(u_0) \cdot R(u_0, u_1) \cdots R(u_{|u|-2}, u_{|u|-1})$$

Dostáváme posloupnost rozdělení $P_{Y_n}(a) = \mathbb{P}[Y_n = a]$ a podle Tvrzení 33, $P_{Y_n} = Q \cdot R^n$ konverguje ke stacionárnímu rozdělení P, které splňuje $P \cdot R = P$. Ukážeme si, že divergence $\mathcal{D}(P_{Y_n}||P)$ je nerostoucí funkce.

Věta 36 *Nechť R je primitivní stochastická matice nad A, $P, Q \in \Delta(A)$ a X, Y odpovídající markovské procesy s počátečním rozděleními $P_{X_0} = P$, $P_{Y_0} = Q$. Pak $\mathcal{D}(P_{Y_n}||P_{X_n})$ je nerostoucí funkce. Speciálně pokud P je stacionární rozdělení příslušné k R, pak $\mathcal{D}(P_{Y_n}||P)$ je nerostoucí posloupnost, která konverguje k nule.*

Důkaz: Podle Lemma 8 platí

$$\begin{aligned}
\mathcal{D}(P_{Y_{n+1}}||P_{X_{n+1}}) &= \sum_{b\in A}\left(\sum_{a\in A} P_{Y_n}(a)\cdot R(a,b)\right)\cdot \log\frac{\sum_{a\in A} P_{Y_n}(a)\cdot R(a,b)}{\sum_{a\in A} P_{X_n}(a)\cdot R(a,b)} \\
&\leq \sum_{b\in A}\sum_{a\in A} \mathcal{P}_{Y_n}(a)\cdot R(a,b)\cdot\log\frac{P_{Y_n}(a)\cdot R(a,b)}{P_{X_n}(a)\cdot R(a,b)} = \mathcal{D}(P_{Y_n}||P_{X_n}) \;\square
\end{aligned}$$

2.3 Markovské aproximace

Definice 15 *Pro stacionární markovský proces řádu $k \geq 1$ definujme* **stacionární rozdělení** $P \in \Delta(A^k)$ *a* **přechodovou matici** $R : A^k \times A \to [0,1]$*:*

$$P(u) = \mathcal{P}_X(u), \quad R(u,a) = \mathcal{P}_X(ua)/\mathcal{P}_X(u), \quad u \in A^k, a \in A$$

Přechodová matice R je typu $A^k \times A$ a stacionární rozdělení je vektor typu $1 \times A^k$. Platí

$$\begin{aligned}
\sum_{a\in A} R(u,a) &= 1, \quad u \in A^k \\
\sum_{a\in A} P(au)\cdot R(au,b) &= P(ub), \quad u \in A^{k-1}, b \in A
\end{aligned}$$

Naopak každý vektor P a matice R která splňuje tyto podmínky určuje stacionární markovský proces řádu k předpisem

$$\mathcal{P}_X(u) = \begin{cases} \sum_{v\in A^{k-|u|}} P(uv) & \text{pro} \quad |u| < k \\ P(u) & \text{pro} \quad |u| = k \\ P(u_{[0,k)}) \cdot R(u_{[0,k)}, u_k) \cdots R(u_{[n-k-1,n-1)}, u_{n-1}) & \text{pro} \quad |u| > k \end{cases}$$

Tvrzení 37 *Entropie stacionárního markovského procesu řádu k je*

$$\begin{aligned}\mathcal{H}(X) &= \mathcal{H}(X_k|X_{[0,k)}) = \mathcal{H}(X_{[0,k]}) - \mathcal{H}(X_{[0,k)}) \\ &= -\sum_{u\in A^k} P(u) \sum_{a\in A} R(u,a)\cdot \log R(u,a)\end{aligned}$$

Důkaz: Pro $n > k$ je $\mathcal{H}(X_n|X_{[0,n)}) = \mathcal{H}(X_n|X_{[n-k,n)}) = \mathcal{H}(X_k|X_{[0,k)})$, takže

$$\begin{aligned}\mathcal{H}(X) &= \mathcal{H}(X_k|X_{[0,k)}) = \sum_{u\in A^k}\sum_{a\in A} \mathcal{P}_X(ua)\cdot \log \frac{\mathcal{P}_X(u)}{\mathcal{P}_X(ua)} \\ &= -\sum_{u\in A^k}\sum_{a\in A} P(u)\cdot R(u,a)\cdot \log R(u,a). \ \square\end{aligned}$$

Definice 16 *Nechť X je stacionární proces. Jeho k-tá* **markovská aproximace** $X^{(k)}$ *je stacionární markovský proces řádu k se stacionárním rozdělením a přechodovou maticí*

$$P(u) = \mathcal{P}_X(u), \ \ R(u,a) = \mathcal{P}_X(ua)/\mathcal{P}_X(u), \ \ u\in A^k, a\in A$$

Entropie stacionárního procesu je tedy limita entropií jeho markovských aproximací

$$\mathcal{H}(X) = \lim_{k\to\infty} \mathcal{H}(X^{(k)}).$$

Markovskou aproximaci lze odhadnout z dat. Je-li $x \in A^N$ text jehož entropii chceme odhadnout a $k < N$, určíme frekvence výskytu slov $u \in A^{k+1}$ v x vzorcem

$$Q(u) = |\{i < N-k : \ x_{[i,i+k]} = u\}|/(N-k), \ \ u\in A^{k+1}.$$

Podmíněné pravděpodobnosti pak jsou $R(u,a) = Q(ua)/\sum_{b\in A} Q(ub)$. K těmto podmíněným pravděpodobnostem pak lze najít stacionární vektor P. Aby tato aproximace byla statisticky spolehlivá, je třeba aby text x byl delší než počet A^{k+1} slov délky k, nebo alespoň delší než počet slov délky $k+1$, které se v něm vyskytují. Tento počet značíme $N(k)$.

Pro anglickou abecedu s 27 písmeny (včetně mezery) je entropie rovnoměrného rozdělení $\log 27 \approx 4.755$. Entropie markovských aproximací anglického textu postupně klesá až k hodnotě asi 1.5. Náhodně generované texty s těmito rozděleními jsou uvedeny níže. Při $k \geq 5$ začnou v textech převládat anglická slova. Délka zkušebního textu [3] zde je $|x| = 196236$

$k = 0$, $N(0) = 27$, $\mathcal{H}(X^{(0)}) = 4.07$
oi re urslrho een iwl tueahueas ia dooaeigehwmis ehtdie yshf rnost taasattnsgitxnece uhiw ed tus ceiiw sro erflranesnangtaa t eaurh r nop eaaeneemhre hotom encnobertag sd h ertgh dg ss phs si egelsf sluh bumwnataesteehdowisulk onehmttvlo edhtehbyaaancd

$k = 1$, $N(1) = 430$, $\mathcal{H}(X^{(1)}) = 3.64$
te igatofofr e rind tin theco ordeconguchino ocliroutorer in buge arourde s cthe e abeisonche ndethedloug ico ke insivalle ct oreid y iot tongnif iorere ts wopr anatot cersur thy initre on an ngive stis peingigstitheme assunll wie goneve bl gn wher deveremathe

$k = 2$, $N(2) = 3033$, $\mathcal{H}(X^{(2)}) = 3.23$
lons againd be whatession th posemon mulatetessibleaducce noted iderectimand of towere bable are pect of rely beiver to causper whinevid laireatur emsenerelse ansights and be the emay inact the thiscreffe ther cander ever been the it whicurety withe it frowered

$k = 3$, $N(3) = 10510$, $\mathcal{H}(X^{(3)}) = 2.87$
thoughout about besistence ope advance ourself is notwithmeter perable natural perstance familinded in conce ideasure or body that the words mar cleas arguing sumstand particater

an the can as it with thered that whence of lain esteneral substractivery whence

$k = 4$, $N(4) = 24807$, $\mathcal{H}(X^{(4)}) = 2.57$
at all have being sect from him which is any reasoning perceive it is neither this i answer of fined to disagrees the affectly how they accounts frame and see sensible and that their sufficulty them were are yet the spokens of the leased laws we have in to impossibility

$k = 5$, $N(5) = 44237$, $\mathcal{H}(X^{(5)}) = 2.33$
ion of part of ordinary help of whence to this opinion the cause of difference of a mean particular and and that solidity and cleared to what any parts in the colours arising our view and irreconcile confusion so high an infinitely and lastly in there is a positions

$k = 6$, $N(6) = 66144$, $\mathcal{H}(X^{(6)}) = 2.12$
ame time apple others are of sensible qualities or is a received opinion that god i answer to be the reasonings should go before are strong lively or abstract ideas that the colour figure merely for the sense and objects of them dependency towards omniscience

$k = 7$, $N(7) = 88421$, $\mathcal{H}(X^{(7)}) = 1.94$
signify and study there be some difficult and defective in the mind is extension is the schoolmen thought strange because they are delivered let him but rather than reason can indifferent from its being by anomalous but only of what kind soever arguments and less

$k = 8$, $N(8) = 108940$, $\mathcal{H}(X^{(8)}) = 1.79$
ttle attentive though we indulgent methods of providence which seems no less plain that without the mind then complaint is grounded on the operating or strictly speaking to be uncreated anew the objected that a couple of children or the yet unexercised mind as

$k = 9$, $N(9) = 126610$, $\mathcal{H}(X^{(9)}) = 1.65$
but all things in terms borrowed from the embarras and delusion of nature whoever considered that i have in view the universal ideas i come now to make some observation and consider them prescinded as well from the ideas of number of parts so great a stress on

$k = 10$, $N(10) = 141125$, $\mathcal{H}(X^{(10)}) = 1.52$
philosophers to employ their thoughts nor passions of virtue and the like arise immediate signification of thought is an idea is evident from what i do i intreat the reader for the existence of matter which argues both the wisdom and goodness in their creator

$k = 11$, $N(11) = 152723$, $\mathcal{H}(X^{(11)}) = 1.41$
eaning you may put them together with the exact harmony and contrivances of the mind perceiving substance it remains therefore to be wondered at if it run into another and more enlightened parts which really they do not excite in us proper sentiments or disagreeable

$k = 12$, $N(12) = 161826$, $\mathcal{H}(X^{(12)}) = 1.32$
e common use of language as is commonly supposed that the particular qualities which combined together and so far is that gravitation towards the moon which to him doth not appear odd or anomalous but only a particular colour wherein all men partake so likewise

Entropii textu můžeme odhadnout ještě jiným způsobem. Uvažujme hru kdy máme postupně odkrývat neznámý text $x \in A^N$. Známe-li ji jeho prefix $x_{[0,n)}$, smíme se ptát zda $x_{[n,n+k)} \in M$. Přitom volíme $k > 0$ a množinu $M \subseteq A^k$. Partner, který text zná, nám na tyto otázky odpovídá. Další úsek $x_{[n+k)}$ zjistíme, pokud na otázku $x_{[n,n+k)} \in \{u\}$, kde $u \in A^k$, dostaneme kladnou odpověď. Je-li $y_i \in B$ odpověď na i-tou otázku, lze $y \in B^*$ považovat za kód textu $x \in A^*$, a jeho délku za entropii textu x. Hraje-li totiž tuto hru počítačový program, lze ze znalosti tohoto programu a ze znalosti y rekonstruovat neznámý text x. Experimenty s touto hrou (s lidskými subjekty) ukazují, že entropie anglického textu se pohybuje okolo hodnoty 1.5. Entropie textu se však může značně lišit v různých jazycích i u různých autorů jak ukazuje srovnání nápisů v tramvajích:

Des chocs imprévisible pouvant se produire, les passagers debout sont priés de se tenir aux mains courants - Brusel 1985

Přidržujte se zadržovacích tyčí - Praha 1985

Bitte festhalten - Berlin 1985

2.4 Zákony velkých čísel

Mnohé zákonitosti teorie informace se vyjadřují v termínech konvergence náhodných veličin a odvozují se z limitních vět teorie pravděpodobnosti. Slabý zákon velkých čísel je založen na konvergenci v pravděpodobnosti. Pro posloupnost reálných náhodných veličin $(X_n)_{n\geq 0}$ a reálnou náhodnou veličinu X píšeme

$$\begin{aligned}
\lim_{n\to\infty} X_n = X \ \textsf{i.p.} \quad &\iff \quad \forall\varepsilon>0,\ \lim_{n\to\infty} \mathbb{P}[|X_n - X| < \varepsilon] = 1\\
\lim_{n\to\infty} X_n \leq X \ \textsf{i.p.} \quad &\iff \quad \forall\varepsilon>0,\ \lim_{n\to\infty} \mathbb{P}[X_n - X < \varepsilon] = 1\\
\lim_{n\to\infty} X_n \geq X \ \textsf{i.p.} \quad &\iff \quad \forall\varepsilon>0,\ \lim_{n\to\infty} \mathbb{P}[X_n - X > -\varepsilon] = 1
\end{aligned}$$

Je-li $\lim_{n\to\infty} X_n = X$ **i.p.** , říkáme, že X_n konvergují k X **v pravděpodobnosti**. Zřejmě

$$\lim_{n\to\infty} X_n = X \ \textsf{i.p.} \iff \lim_{n\to\infty} X_n \leq X \ \textsf{i.p.} \ \text{a} \ \lim_{n\to\infty} X_n \geq X \ \textsf{i.p.}$$

Slabý zákon velkých čísel plyne z Čebyševovy nerovnosti.

Tvrzení 38 (Čebyševova nerovnost) *Nechť X je reálná náhodná veličina s konečnou střední hodnotou $E = \mathbb{E}(X)$ a konečným rozptylem $D = \mathbb{V}(X)$. Pak pro každé $\varepsilon > 0$ platí*

$$\mathbb{P}[|X - E| \geq \varepsilon] \leq D/\varepsilon^2.$$

Důkaz: Položme $A_1 = \{a \in A : \ |a - E| \geq \varepsilon\}$. Pak

$$D \ \geq \ \sum_{a\in A_1} P_X(a)\cdot|a-E|^2 \geq \sum_{a\in A_1} P_X(a)\cdot\varepsilon^2 = \varepsilon^2\cdot\mathbb{P}[X\in A_1] \ \square$$

Pro násobek náhodné veličiny platí $\mathbb{E}(aX) = a\cdot\mathbb{E}(X)$ a $\mathbb{V}(aX) = a^2\cdot\mathbb{V}(X)$. Pro součet $Z_n = X_0 + \cdots + X_{n-1}$ náhodných veličin platí $\mathbb{E}(Z_n) = \mathbb{E}(X_0) + \cdots + \mathbb{E}(X_{n-1})$. Jsou-li X_i nezávislé, je $\mathbb{V}(Z_n) = \mathbb{V}(X_0) + \cdots + \mathbb{V}(X_{n-1})$.

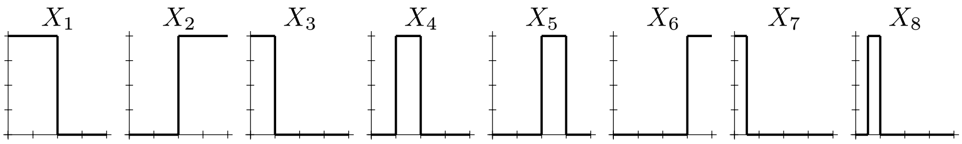
Věta 39 (Slabý zákon velkých čísel) *Nechť X_n je posloupnost nezávislých reálných náhodných veličin se stejnou střední hodnotou $\mathbb{E}(X_n) = E < \infty$ a stejným rozptylem $\mathbb{V}(X_n) < \infty$. Pak*

$$\lim_{n\to\infty} \frac{X_0 + \cdots + X_{n-1}}{n} = E \ \textsf{i.p.}$$

Důkaz: Označme $Z_n = (X_0 + \cdots + X_{n-1})/n$. Pak $\mathbb{E}(Z_n) = E$, $\mathbb{V}(Z_n) = D/n$. Podle Čebyševovy nerovnosti platí $\mathbb{P}[|Z_n - E| \geq \varepsilon] \leq D/n\varepsilon^2$. $\square$

Říkáme že k pravděpodobnostní jev $M \subseteq \Omega$ je skoro jistý, pokud jeho pravděpodobnost je 1. Říkáme že posloupnost náhodných veličin $(X_n)_{n\in\mathbb{N}}$ konverguje k náhodné veličině X skoro jistě, a píšeme $\lim_{n\to\infty} X_n = X$ **s.j.** pokud $\mathbb{P}[\omega\in\Omega : \ \lim_{n\to\infty} X_i(\omega) = X(\omega)] = 1$. Z konvergence skoro jistě vyplývá konvergence v pravděpodobnosti, naopak to však neplatí.

Příklad 9 *Existuje posloupnost náhodných veličin $(X_i : \Omega \to [0,1])_{i\geq 0}$ taková, že $\lim_{i\to\infty} X_i = 0$ **i.p.** ale neplatí $\lim_{i\to\infty} X_i = 0$ **s.j.***



Obrázek 12: Konvergence v pravděpodobnosti a skoro jistě

Důkaz: Na pravděpodobnostním prostoru $\Omega = [0, 1]$ s lebesgueovskou pravděpodobnostní mírou uvažujme náhodné veličiny $X_n : \Omega \to \{0, 1\}$ (viz obrázek 12) definované předpisem

$$X_n(\omega) = \begin{cases} 1 & \text{pokud } \frac{n+1-2^p}{2^p} \le \omega < \frac{n+2-2^p}{2^p}, \text{ kde } p = \lfloor \log(n+1) \rfloor \\ 0 & \text{jinak} \end{cases}$$

Pak $\mathbb{P}[X_n \neq 0] = 2^{-p}$, takže $\lim_{n\to\infty} X_n = 0$ **i.p.** Na druhé straně pro každé $\omega \in \Omega$ platí $\limsup_{n\to\infty} X_n(\omega) = 1$, takže neplatí $\lim_{n\to\infty} X_n = 0$ **s.j.** □

Konvergenci skoro jistě lze však přesto charakterizovat konvergencí v pravděpodobnosti pomocí suprema:

Tvrzení 40 *Nechť X a $(X_n)_{n\ge 0}$ jsou reálné náhodné veličiny.*

$$\begin{aligned} \lim_{n\to\infty} \sup_{m\ge n} X_m \le X \text{ i.p.} \quad &\iff \quad \limsup_{n\to\infty} X_n \le X \text{ s.j.} \\ \lim_{n\to\infty} \inf_{m\ge n} X_m \ge X \text{ i.p.} \quad &\iff \quad \liminf_{n\to\infty} X_n \ge X \text{ s.j.} \\ \lim_{n\to\infty} \sup_{m\ge n} |X_m - X| = 0 \text{ i.p.} \quad &\iff \quad \lim_{n\to\infty} X_n = X \text{ s.j.} \end{aligned}$$

Důkaz: Stačí dokázat první ekvivalenci, druhé dvě z ní plynou. Pro $\varepsilon > 0$ a $n \in \mathbb{N}$ položme

$$M_{\varepsilon,n} = \{\omega \in \Omega : \sup_{m\ge n} X_m(\omega) < X(\omega) + \varepsilon\}, \quad M_\varepsilon = \bigcup_{n\in\mathbb{N}} M_{\varepsilon,n}, \quad M = \bigcap_{k>0} M_{\frac{1}{k}}$$

Nechť $\lim_{n\to\infty} \sup_{m\ge n} X_m \le X$ **i.p.** Pak pro každé $\varepsilon > 0$ a každé $\delta > 0$ existuje $n > 0$ takové že $\mathbb{P}(M_{\varepsilon,n}) \ge 1-\delta$, takže $\mathbb{P}(M_\varepsilon) > 1-\delta$. To znamená že $\mathbb{P}(M_\varepsilon) = 1$ a $\mathbb{P}(M) = 1$. Je-li $\omega \in M$, pak pro každé k je $\limsup_{n\to\infty} X_n(\omega) \le X(\omega) + \frac{1}{k}$, takže $\limsup_{n\to\infty} X_n(\omega) \le X(\omega)$. To znamená $\limsup_{n\to\infty} X_n \le X$ **s.j.**

Naopak předpokládejme že $\limsup_{n\to\infty} X_n \le X$ **s.j.** Je-li $\limsup_{n\to\infty} X_n(\omega) \le X(\omega)$, a $\varepsilon > 0$, je $\omega \in M_\varepsilon$, takže $\mathbb{P}(M_\varepsilon) = 1$. Protože $M_{\varepsilon,n} \subseteq M_{\varepsilon,n+1}$, pro každé $\delta > 0$ existuje n takové, že $\mathbb{P}(M_{\varepsilon,n}) > 1 - \delta$. Z toho plyne $\lim_{n\to\infty} \sup_{m\ge n} X_m \le X$ **i.p.** □

Z Tvrzení 40 bezprostředně plyne

Tvrzení 41 *Nechť X a $(X_n)_{n\ge 0}$ jsou reálné náhodné veličiny. Pak*

$$\begin{aligned} \limsup_{n\to\infty} X_n \le X \text{ s.j.} \quad &\implies \quad \lim_{n\to\infty} X_n \le X \text{ i.p.} \\ \liminf_{n\to\infty} X_n \ge X \text{ s.j.} \quad &\implies \quad \lim_{n\to\infty} X_n \ge X \text{ i.p.} \\ \lim_{n\to\infty} X_n = X \text{ s.j.} \quad &\implies \quad \lim_{n\to\infty} X_n = X \text{ i.p.} \end{aligned}$$

Věta 42 (Silný zákon velkých čísel) *Nechť $(X_i)_{i\ge 0}$ je posloupnost nezávislých reálných náhodných veličin se stejným rozdělením se střední hodnotou $\mathbb{E}(X_i) = E < \infty$ a konečným rozptylem. Pak*

$$\lim_{n\to\infty} \frac{X_0 + \cdots + X_{n-1}}{n} \to E \text{ s.j.}$$

Důkaz viz Rényi [25].

Definice 17 **Frekvenční rozdělení** $\mathfrak{P}_u \in \Delta(A)$ *slova* $u \in A^*$ *je*

$$\mathfrak{P}_u(a) = |u|_a/|u|, \quad kde\ |u|_a = \{i < |u| : \ u_i = a\}$$

Věta 43 *Nechť* $(X_i : \Omega \to A)_{i\in\mathbb{N}}$ *je bernoulliovský proces s rozdělením* $P \in \Delta(A)$. *Pak*

$$\lim_{n\to\infty} \mathcal{H}(\mathfrak{P}_{X_{[0,n)}}) = \mathcal{H}(P) \text{ s.j.}$$

Důkaz: Platí $|X_{[0,n)}|_a = \sum_{i<n} |X_i|_a/n$. Protože $\mathbb{E}(|X_i|_a) = P(a)$, podle silného zákonu velkých čísel $\mathfrak{P}_{X_{[0,n)}}(a) = |X_{[0,n)}|_a/n$ konverguje skoro jistě k $P(u)$. Z toho plyne že $\mathcal{H}(\mathfrak{P}_{X_{[0,n)}})$ konverguje skoro jistě k $\mathcal{H}(P)$. □

Věta 44 (Shannon) *Nechť* $(X_i : \Omega \to A)_{i\in\mathbb{N}}$ *je bernoulliovský proces s rozdělením* $P \in \Delta(A)$. *Pak*

$$\lim_{n\to\infty} \frac{\mathfrak{I}_{X_{[0,n)}}}{n} = \mathcal{H}(P) \text{ s.j.}$$

Důkaz: Položme $Y_n = \mathfrak{I}_{X_{[0,n)}}/n = \sum_{i<n} \mathfrak{I}_{X_i}/n$. Protože $\mathfrak{I}_{X_i}$ jsou navzájem nezávislé stejně rozdělené veličiny se střední hodnotou $\mathbb{E}(\mathfrak{I}_{X_i}) = \mathcal{H}(X_i) = \mathcal{H}(P)$, je $\lim_{n\to\infty} Y_n = \mathbb{E}(\mathfrak{I}_{X_0}) = \mathcal{H}(P)$ s.j. □

Věta 45 (o rovnoměrném rozdělení) *Nechť* $(X_i : \Omega \to A)_{i\in\mathbb{N}}$ *je bernoulliovský proces s rozdělením* $P \in \Delta(A)$. *Pak* $\lim_{n\to\infty} \mathfrak{I}_{X_{[0,n)}}/n = \mathcal{H}(P)$ i.p. , *tj. pro každé* $\varepsilon > 0$ *platí*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\mathcal{H}(P) + \frac{1}{n}\log\prod_{i=0}^{n-1} P(X_i)\right| < \varepsilon\right] = 1$$

Důkaz bezprostředně plyne z Věty 44. Odhadneme pravděpodobnost "typické množiny" nejpravděpodobnějších slov. Připomeňme, že pro $P \in \Delta(A)$ je $P^n \in \Delta(A^n)$ definováno jako součin pravděpodobností písmen, a pro $M \subseteq A^n$ je $P^n(M) = \sum_{u\in M} P^n(u)$.

Definice 18 *Pro rozdělení* $P \in \Delta(A)$, $\varepsilon > 0$ *a* $n > 0$ *je* **typická množina** $\mathcal{M}^n_\varepsilon(P)$ *definována vzorcem*

$$\begin{aligned}\mathcal{M}^n_\varepsilon(P) &= \left\{u \in A^n : \left|\mathcal{H}(P) + \frac{1}{n}\log\prod_{i=0}^{n-1} P(u_i)\right| < \varepsilon\right\} \\ &= \left\{u \in A^n : \ 2^{-n(\mathcal{H}(P)+\varepsilon)} < P^n(u) < 2^{-n(\mathcal{H}(P)-\varepsilon)}\right\}\end{aligned}$$

Do typicke množiny frekvenčního rozdělení $\mathfrak{P}_u$ patří speciálně samo slovo u:

Tvrzení 46 *Pro* $u \in A^*$ *je* $\mathfrak{P}^{|u|}_u(u) = 2^{-n\cdot\mathcal{H}(\mathfrak{P}_u)}$.

Důkaz:

$$\prod_{i<|u|} \mathfrak{P}_u(u_i) = \prod_{a\in A} \mathfrak{P}_u(a)^{|u|_a} = \prod_{a\in A} 2^{|u|\cdot\mathfrak{P}_u(a)\cdot\log\mathfrak{P}_u(a)} = 2^{-|u|\cdot\mathcal{H}(\mathfrak{P}_u)} \quad \square$$

Věta 47 (o typické množině) *Pro každé* $\delta, \varepsilon > 0$ *a pro každé dosti velké* $n > n_{\delta,\varepsilon}$ *platí*

$$(1-\delta)\cdot 2^{n\cdot(\mathcal{H}(P)-\varepsilon)} \le |\mathcal{M}^n_\varepsilon(P)| \le 2^{n\cdot(\mathcal{H}(P)+\varepsilon)}, \quad P^n(\mathcal{M}^n_\varepsilon(P)) > 1-\delta$$

Důkaz: Nechť X je bernoulliovský proces s rozložením P. Z Věty 45 o rovnoměrném rozdělení bezprostředně plyne $\mathbb{P}[X_{[0,n)} \in \mathcal{M}_\varepsilon^n(P)] > 1-\delta$ pro všechna dosti velká n. Dále

$$1 \geq \sum_{u \in \mathcal{M}_\varepsilon^n(P)} P^n(u) \geq |\mathcal{M}_\varepsilon^n(P)| \cdot 2^{-n(\mathcal{H}(P)+\varepsilon)}$$

$$1-\delta \leq \sum_{u \in \mathcal{M}_\varepsilon^n(P)} P^n(u) \leq |\mathcal{M}_\varepsilon^n(P)| \cdot 2^{-n(\mathcal{H}(P)-\varepsilon)}. \quad \square$$

2.5 Teorie typů

Frekvenční vektor $\mathfrak{P}_u$ sestává z racionálních čísel $\mathfrak{P}_u(a) = |u|_a/|u|$, jejichž jmenovatel je $|u|$. Prostor takových vektorů značíme

$$\Delta^n(A) = \{P \in \Delta(A) : \ \forall a \in A, \ n \cdot P(a) \in \mathbb{N}\}$$

Pro $P \in \Delta^n(A)$ položme

$$\mathcal{C}^n(P) = \{u \in A^n : \ \mathfrak{P}_u = P\}, \ \text{ takže } \ |\mathcal{C}^n(P)| = n!/\prod_{a \in A}(n \cdot P(a))!$$

Podobně pro $u \in A^*$ položme

$$\mathcal{C}(u) = \{v \in A^{|u|} : \ \mathfrak{P}_v = \mathfrak{P}_u\}, \ \text{ takže } \ |\mathcal{C}(u)| = |u|!/\prod_{a \in A}|u|_a!$$

Zobecníme nyní Tvrzení 46.

Věta 48 *Nechť $Q \in \Delta(A)$ a $u \in A^n$. Pak $\mathfrak{I}_{Q^n}(u) = |u| \cdot (\mathcal{H}(\mathfrak{P}_u) + \mathcal{D}(\mathfrak{P}_u||Q))$.*

Důkaz:

$$Q^n(u) = \prod_{i<|u|} Q(u_i) = \prod_{a \in A} Q(a)^{|u|_a} = \prod_{a \in A} 2^{|u| \cdot \mathfrak{P}_u(a) \cdot \log Q(a)} = 2^{-|u| \cdot (\mathcal{H}(\mathfrak{P}_u) + \mathcal{D}(\mathfrak{P}_u||Q))},$$

takže $\mathfrak{I}_{Q^n}(u) = -\log Q^n(u) = |u| \cdot (\mathcal{H}(\mathfrak{P}_u) + \mathcal{D}(\mathfrak{P}_u||Q))$. $\square$

Tvrzení 49 $|\Delta^n(A)| \leq (n+1)^{|A|-1}$.

Důkaz: Pro každé $P(a) \in \{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$ máme nejvýše $n+1$ možností, přitom poslední složka je určena předcházejícími. $\square$

Věta 50 *Nechť $P \in \Delta^n(A)$, $u \in A^n$. Pak*

$$\frac{2^{n \cdot \mathcal{H}(P)}}{(n+1)^{|A|-1}} \leq |\mathcal{C}^n(P)| \leq 2^{n \cdot \mathcal{H}(P)}, \quad \frac{2^{n \cdot \mathcal{H}(\mathfrak{P}_u)}}{(n+1)^{|A|-1}} \leq |\mathcal{C}(u)| \leq 2^{n \cdot \mathcal{H}(\mathfrak{P}_u)}$$

Důkaz: Podle Tvrzení 48

$$1 \geq P^n(\mathcal{C}^n(P)) = \sum_{u \in \mathcal{C}^n(P)} P^n(u) = \sum_{u \in \mathcal{C}^n(P)} 2^{-n \cdot \mathcal{H}(P)} = |\mathcal{C}_n(P)| \cdot 2^{-n \cdot \mathcal{H}(P)}$$

takže $|\mathcal{C}^n(P)| \leq 2^{n \cdot \mathcal{H}(P)}$. Pro opačnou nerovnost použijeme vztah $m! \geq n! \cdot n^{m-n}$, který lze dokázat rozborem případů $m \geq n$ a $m < n$. Pro $P, Q \in \Delta^n(A)$ platí

$$\begin{aligned}
\frac{P^n(\mathcal{C}_n(P))}{P^n(\mathcal{C}_n(Q))} &= \frac{|\mathcal{C}_n(P)| \cdot \prod_{a \in A} P(a)^{nP(a)}}{|\mathcal{C}_n(Q)| \cdot \prod_{a \in A} P(a)^{nQ(a)}} = \prod_{a \in A} \frac{(nQ(a))!}{(nP(a))!} P(a)^{nP(a)-nQ(a)} \\
&\geq \prod_{a \in A} (nP(a))^{nQ(a)-nP(a)} \cdot P(a)^{nP(a)-nQ(a)} = \prod_{a \in A} n^{n(Q(a)-P(a))} \\
&= n^{n \cdot \sum_{a \in A}(Q(a)-P(a))} = n^0 = 1
\end{aligned}$$

Odtud z Tvrzení 49 a 48 plyne

$$\begin{aligned} 1 &= \sum_{Q \in \Delta^n(A)} P^n(\mathcal{C}^n(Q)) \le \sum_{Q \in \Delta^n(A)} P^n(\mathcal{C}^n(P)) \le (n+1)^{|A|-1} \cdot P^n(\mathcal{C}^n(P)) \\ &= (n+1)^{|A|-1} \cdot 2^{-n \cdot \mathcal{H}(P)} \cdot |\mathcal{C}^n(P)| \quad \square \end{aligned}$$

Pro binární abecedu lze tvrzení zesílit

Tvrzení 51 *Pro $k \le n/2$ platí*

$$\sum_{j=0}^{k} \binom{n}{j} \le 2^{n \cdot \mathcal{H}(\frac{k}{n}, \frac{n-k}{n})}$$

Důkaz:

$$\begin{aligned} 1 &\ge \sum_{j=0}^{k} \binom{n}{j} \cdot \left(\frac{k}{n}\right)^j \cdot \left(\frac{n-k}{n}\right)^{n-j} \ge \sum_{j=0}^{k} \binom{n}{j} \cdot \left(\frac{k}{n}\right)^k \cdot \left(\frac{n-k}{n}\right)^{n-k} \\ &\ge \sum_{j=0}^{k} \binom{n}{j} \cdot 2^{-n \cdot \mathcal{H}(\frac{k}{n}, \frac{n-k}{n})} \quad \square \end{aligned}$$

Věta 52 *Pro každé $P \in \Delta^n(A)$ a $Q \in \Delta(A)$ platí*

$$\frac{1}{(n+1)^{|A|-1}} 2^{-n \cdot \mathcal{D}(P||Q)} \le Q^n(\mathcal{C}^n(P)) \le 2^{-n \cdot \mathcal{D}(P||Q)}$$

a tedy

$$\lim_{k \to \infty} \frac{-\log Q^{nk}(\mathcal{C}^{nk}(P))}{nk} = \mathcal{D}(P||Q)$$

Důkaz: Podle Tvrzení 48 platí

$$Q^n(\mathcal{C}^n(P)) = \sum_{u \in \mathcal{C}^n(P)} Q^n(u) = \sum_{u \in \mathcal{C}^n(P)} 2^{-n \cdot (\mathcal{D}(P||Q) + \mathcal{H}(P))} = |\mathcal{C}^n(P)| \cdot 2^{-n \cdot (\mathcal{D}(P||Q) + \mathcal{H}(P))}$$

Věta plyne dosazením tohoto vztahu do nerovnosti z Tvrzení 50. $\square$

Speciálně pro $P = Q$ dostáváme $\lim_{k \to \infty} -\log P^n(\mathcal{C}^n(P))/n = \mathcal{D}(P||P) = 0$. Ze Stirlingovy formule dostáváme přesnější odhad. Pro každé k které je dělitelné n platí

$$P^k(\mathcal{C}^k(P)) = k! \cdot \prod_{a \in A} \frac{P(a)^{kP(a)}}{(kP(a))!} \approx k^{-\frac{|A|-1}{2}}$$

Tato pravděpodobnost sice konverguje k nule, konvergence je však polynomiální a nikoliv exponenciální.

3 Komprese dat

3.1 Blokové kódy

Blokový kód abecedy A v binární abecedě $B = \{0, 1\}$ je každé zobrazení $f : A \to B^+$. Blokový kód rozšiřujeme na zobrazení $f : A^* \to B^*$ a $f : A^{\mathbb{N}} \to B^{\mathbb{N}}$ konkatenací $f(u_0 u_1 u_2 \ldots) = f(u_0)f(u_1)f(u_2)\ldots$.

Definice 19 *Nechť $f : A \to B^+$ je blokový kód.*

(1) Kód f je **prefixový**, *pokud pro žádná $a \neq a' \in A$ neplatí $f(a) \sqsubseteq_p f(a')$.*

(2) Kód f je **jednoznačný**, *pokud $f : A^* \to B^*$ je prosté zobrazení.*

(3) Kód f je **úplný**, *pokud $f : A^{\mathbb{N}} \to A^{\mathbb{N}}$ je surjektivní zobrazení.*

Kód $a \mapsto 0$, $b \mapsto 10$, $c \mapsto 110$, $d \mapsto 111$ je prefixový, jednoznačný a úplný. Omezíme-li ho na abecedu $\{a, b, c\}$, dostaneme prefixový a jednoznačný kód, který není úplný. Kód $a \mapsto 0$, $b \mapsto 01$, $c \mapsto 011$, $d \mapsto 111$ je jednoznačný (sufixový) kód který není ani prefixový ani úplný.

Tvrzení 53 *Každý prefixový kód je jednoznačný.*

Důkaz: Pokud $f(u) = f(v)$, pak buď $f(u_0)$ je prefix $f(v_0)$ nebo naopak. V každém případě je $u_0 = v_0$ a indukcí dostaneme $u = v$. $\square$

Věta 54 (Kraftova nerovnost) *Nechť přirozená čísla $d_0, \ldots, d_{k-1}$ splňují nerovnost*

$$2^{-d_0} + \cdots + 2^{-d_{k-1}} \leq 1.$$

Pak existuje prefixový kód $f : \mathbb{Z}_k \to B^+$ s délkami $|f(i)| = d_i$.

Důkaz: Seřadíme délky podle velikosti $1 \leq d_0 \leq d_1 \leq \cdots \leq d_{k-1}$. Zvolme libovolně kód $f(0) \in B^{d_0}$ s délkou d_0. Pak pro množinu slov délky d_1, jejich prefix je $f(0)$ platí

$$|\{u \in B^{d_1} : f(0) \sqsubseteq_p u\}| = 2^{d_1 - d_0} < 2^{d_1 - d_0} + 1 = 2^{d_1}(2^{-d_0} + 2^{-d_1}) \leq 2^{d_1} = |B^{d_1}|$$

Existuje tedy slovo $f(1) \in B^{d_1}$ jehož prefix není $f(0)$. Déle postupujeme indukcí. Předpokládejme, že jsme již sestrojili $f(0), \ldots, f(i-1)$, které tvoří prefixový kód. Pak množina slov délky d_i, jejichž prefix je nějaké $f(j)$, kde $j < i$ je disjunktní sjednocení

$$\begin{aligned}
\{u \in B^{d_i} : f(0) \sqsubseteq_p u \vee \cdots \vee f(i-1) \sqsubseteq_p u\} &= \bigcup_{j<i} \{u \in B^{d_i} : f(j) \sqsubseteq_p u\} \\
|\{u \in B^{d_i} : f(0) \sqsubseteq_p u \vee \cdots \vee f(i-1) \sqsubseteq_p u\}| &= 2^{d_i - d_0} + \cdots + 2^{d_i - d_{i-1}} \\
&< 2^{d_i}(2^{-d_0} + \cdots + 2^{-d_{i-1}} + 2^{-d_i}) \leq 2^{d_i}
\end{aligned}$$

takže existuje $f(i) \in B^{d_i}$ jehož prefixem není žádné $f(j)$, kde $j < i$. $\square$

Věta 55 (MacMillan) *Nechť $f : \mathbb{Z}_k \to B^+$ je jednoznačný kód, a $d_i = |f(i)|$. Pak*

$$2^{-d_0} + \cdots + 2^{-d_{k-1}} \leq 1.$$

Důkaz: Předpokládejme opět $1 \leq d_0 \leq \cdots \leq d_{k-1}$ a položme $c = \sum_{i=0}^{k-1} 2^{-d_i}$. Pak $c^2 = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} 2^{-d_i} \cdot 2^{-d_j}$ a pro $p \geq 2$ přirozené je

$$c^p = \sum_{i_0=0}^{k-1} \cdots \sum_{i_{p-1}=0}^{k-1} 2^{-(d_{i_0} + \cdots + d_{i_{p-1}})} = \sum_{j=1}^{pd_{k-1}} c_j 2^{-j}.$$

Pro koeficienty c_j platí

$$\begin{aligned} c_j &= |\{(i_0,\dots,i_{p-1}) \in A^p :\ d_{i_0} + \cdots + d_{i_{p-1}} = j\}| \\ &= |\{(i_0,\dots,i_{p-1}) \in A^p :\ |f(i_0,\dots,i_{p-1})| = j\}|. \end{aligned}$$

Slova $i = (i_0,\dots,i_{p-1}) \in A^p$ jsou zprávy v abecedě A a c_j je počet všech zpráv jejichž kódy mají délku právě j. Protože počet všech možných kódů délky j je 2^j a protože f je jednoznačný, platí $c_j \le 2^j$. Odtud dostáváme

$$c^p \le \sum_{j=1}^{pd_{k-1}} 2^j \cdot 2^{-j} = pd_{k-1} \implies \frac{c^p}{p} \le d_{k-1}.$$

To je možné pouze pokud $c \le 1$. □

3.2 Délka kódu

Definice 20 **Délka blokového kódu** $f : A \to B^+$ *při rozdělení* $P \in \Delta(A)$ *je*

$$L(P,f) = \sum_{a \in A} P(a) \cdot |f(a)|.$$

Délka rozdělení $P \in \Delta(A)$ *je* $L(P) = \min\{L(P,f) :\ f : A \to B^+$ *je jednoznačný kód*$\}$. *Jednoznačný kód* $f : A \to 2^+$ *je* **minimální kód** *pro* P*, pokud* $L(P,f) = L(P)$.

Tvrzení 56 *Nechť* X *je bernoulliovský proces nad* A *a* $f : A \to B^+$ *je blokový kód. Pak*

$$\lim_{n\to\infty} \frac{|f(X_{[0,n)})|}{n} = L(P,f) \text{ s.j.}$$

Důkaz: Platí $|f(X_{[0,n)})| = \sum_{i<n} |f(X_i)|$ a $\mathbb{E}|f(X_i)| = \sum_{a\in A} P(a) \cdot |f(a)| = L(P,f)$. □

Věta 57 (Shannonův kód) *Pro každé rozdělení* $P \in \Delta(A)$ *platí*

$$\mathcal{H}(P) \le L(P) < \mathcal{H}(P) + 1.$$

Důkaz: Pro dané P sestrojíme kód f. Položme $d_a = \lceil -\log P(a) \rceil$, takže

$$\log \frac{1}{P(a)} \le d_a < \log \frac{1}{P(a)} + 1$$

Odtud $2^{-d_a} \le P(a)$ takže čísla d_a splňují Kraftovu nerovnost. Existuje tedy prefixový kód $f : A \to B^+$ pro který $|f(a)| = d_a$. Odtud

$$L(P,f) = \sum_{a\in A} P(a) \cdot d_a < \sum_{a\in A} P(a) \left(\log \frac{1}{P(a)} + 1 \right) = \mathcal{H}(P) + 1.$$

Ukázali jsme tedy $L(P) \le L(P,f) \le \mathcal{H}(P) + 1$. Pro důkaz opačné nerovnosti potřebujeme nerovnost $\ln x \le x - 1$ z Lemma 5. Předpokládejme tedy že $f : A \to B^+$ je jednoznačný kód, takže $d_a = |f(a)|$ splňují Kraftovu nerovnost. Odtud dostáváme

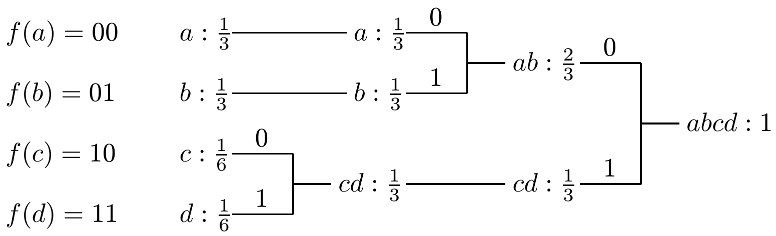
$$\begin{aligned} \mathcal{H}(P) - L(P,f) &= \sum_{a\in A} P(a) \left(\log \frac{1}{P(a)} - \log 2^{d_a} \right) = \frac{1}{\ln 2} \sum_{a\in A} P(a) \ln \frac{2^{-d_a}}{P(a)} \\ &\le \frac{1}{\ln 2} \sum_{a\in A} P(a) \left(\frac{2^{-d_a}}{P(a)} - 1 \right) \le 0 \end{aligned}$$

Pro každý jednoznačný kód je tedy $\mathcal{H}(P) \leq L(P, f)$, takže $\mathcal{H}(P) \leq L(P)$. $\square$

Z důkazu věty plyne, že pro rozdělení $P \in \Delta(P)$, jehož hodnoty jsou záporné mocniny dvou, tj. $P(a) = 2^{-d_a}$ existuje prefixový kód f, jehož délka $L(P, f) = \mathcal{H}(P)$ je přesně jeho entropie. Pokud pravděpodobnosti nejsou mocniny dvou, lze efektivitu kódování zvýšit, kódujeme-li zprávy po blocích. Pro dané rozdělení $P \in \Delta(A)$ a jeho m-tou mocninu $P^m \in \Delta(A^m)$ platí $\mathcal{H}(P^m) = m \cdot \mathcal{H}(P)$. Podle Věty 57 je tedy $m \cdot \mathcal{H}(P) \leq L(P^m) \leq m \cdot \mathcal{H}(P) + 1$, takže $\lim_{m\to\infty} L(P^m)/m = \mathcal{H}(P)$. Při kódování po blocích připadá asymptoticky $\mathcal{H}(P)$ bitů na jedno písmeno. V odstavci 3.5 tento výsledek ještě zobecníme.

3.3 Huffmannův kód

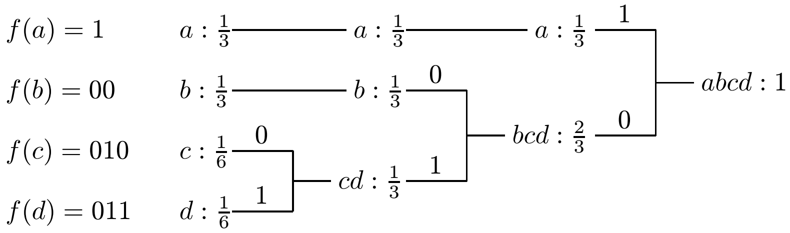
Věty 54 a 57 (přesněji řečeno jejich důkazy) dávají algoritmus jak k danému rozdělení $P \in \Delta(A)$ sestrojit binární prefixový kód $f : A \to B^+$ pro který $L(P, f) < \mathcal{H}(P) + 1$. Tento kód ale nemusí být minimální kód pro P.



Obrázek 13: Huffmannův kód

Příklad 10 *Rozdělení* $P = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$ *má entropii* $\mathcal{H}(P) = \log(3) + \frac{1}{3} \approx 1.918$.

Záporné logaritmy pravděpodobností jsou $-\log P = (1.58, 1.58, 2.58, 2.58)$, takže délky jsou $d = (2, 2, 3, 3)$. Tomu odpovídá například prefixový kód $g = (00, 01, 100, 101)$ s délkou $L(P, g) = \frac{7}{3} \approx 2.333$. Minimální kód pro P je ale například $f = (00, 01, 10, 11)$ s délkou $L(P, f) = 2 < L(P, g)$. Minimální prefixový kód daného rozdělení lze získat Huffmannovým algoritmem. Algoritmus sloučí dvě písmena s nejmenší pravděpodobností do jediného nového písmene a získá tak rozdělení na menší abecedě. Potom sestrojí minimální kód pro tuto menší abecedu a z kódu složeného písmene získá kódy jeho dvou písmen tak že k němu přidáme jeden rozlišovací bit. Postup je znázorněn na obrázku 13. Minimální kód není určen jednoznačně. V případě že v průběhu algoritmu dojde k situaci že dvojice písmen s nejmenší pravděpodobností není určena jednoznačně, vedou obě alternativy k minimálnímu kódu. K této situaci dochází v příkladě 10. Alternativní minimální kód (se stejnou délkou $L(P, f) = 2$) je sestrojen na obrázku 14.



Obrázek 14: Alternativní Huffmannův kód

Definice 21

(1) **Binární strom** *je množina $T \subseteq B^*$ pro kterou platí $u \in T \;\&\; v \sqsubseteq_p u \implies v \in T$.*

(2) **Listy** *binárního stromu T tvoří množinu $L(T) = \{u \in T : u0 \notin T \;\&\; u1 \notin T\}$.*

(3) Binární strom T je **úplný**, *pokud platí $u \in T \implies (u0 \in T \iff u1 \in T)$.*

(4) Binární strom prefixového kódu $f : A \to B^+$ je $T_f = \{u \in B^* : \exists a \in A, u \sqsubseteq_p f(a)\}$.

Tvrzení 58 *Je-li T binární strom a $f : A \to L(T)$ prosté zobrazení, je f prefixový kód. Přitom f je úplný kód právě když T je úplný binární strom a $f : A \to L(T)$ je bijektivní zobrazení.*

Důkaz: Je-li $f : A \to L(T)$ prosté a $f(a)$ vlastní prefix $f(b)$, pak $f(a)$ nemůže být list a to je spor. Nechť f je úplný kód. Jestliže f není bijektivní nebo T není úplný, existuje $a \in A$ takové že buď $f(a) = u0$ a $u1 \notin f[A]$, nebo $f(a) = u1$ a $u0 \notin f[A]$. Pak buď $u0^\infty$ nebo $u1^\infty$ nenáleží do $f[A^{\mathbb{N}}]$, takže f není úplný kód. Naopak předpokládejme, že f je bijekce a T je úplný strom. Je-li u nejdelší prefix $y \in B^{\mathbb{N}}$ který náleží do T, pak $u \in L(T)$. Pokud by totiž u nenáleželo do $L(T)$, pak by obě slova $u0$ a $u1$ náležela do T a jedno z nich by bylo prefixem y. existuje tedy $x_0 \in A$ a $z \in B^{\mathbb{N}}$ takové že $y = f(x_0)z$. Opakováním tohoto postupu sestrojíme $x \in A^{\mathbb{N}}$ pro které $y = f(x)$. $\square$

Tvrzení 59 *Pro každý $P \in \Delta(A)$ existuje minimální prefixový kód. Je-li $f : A \to B^+$ minimální prefixový kód pro P, je T_f úplný binární strom a platí*

$$P(a) < P(b) \implies |f(a)| \geq |f(b)|.$$

Dále existují $a \neq b \in A$ takové, že pro každé $c \in A$ platí $|f(c)| \leq |f(a)| = |f(b)|$ a $f(a)$ se od $f(b)$ liší jen posledním bitem.

Důkaz: Množina kódů $g : A \to B^+$ takových že $L(P, g) \leq \lceil \log |A| \rceil$ je konečná a neprázdná, takže existuje minimální kód g. Podle Věty 55, délky $|g(a)|$ splňují Kraftovu nerovnost a podle Věty 54 existuje prefixový kód f se stejnými délkami. $|f(a)| = |g(a)|$, takže f je také minimální. Předpokládejme sporem, že T_f není minimální strom. Pak existuje $u \in T_f$ takový že $u0 \in T_f$ a $u1 \notin T_f$ (nebo naopak). Sestrojme kód

$$g(v) = \begin{cases} f(v) & \text{pokud} \quad u \not\sqsubseteq_p f(v) \\ uw & \text{pokud} \quad f(v) = u0w \end{cases}$$

Pak g je prefixový kód kratší než f a to je spor, takže T_f je minimální strom. Kdyby pro $P(a) < P(b)$ platilo $|f(a)| < |f(b)|$, pak by záměnou $f(a)$ a $f(b)$ vznikl kratší kód. Nakonec dokažme, že dvě nejdelší kódová slova se liší jen posledním bitem. Nechť $f(a)$ je nejdelší slovo kódu, tj. $|f(c)| \leq |f(a)|$ pro každé $c \in A$. Položme $f(a) = ud$, kde $d \in B$ a uvažujme kód $g : A \to B^*$ definovaný předpisem $g(a) = u$, $g(c) = f(c)$ pro $c \neq a$. Protože f je minimální, g není prefixový a existuje $b \neq a$ pro které $u \sqsubseteq_p f(b)$ nebo $f(b) \sqsubseteq_p u$. Protože f je prefixový kód, není $f(b) \sqsubseteq u$, a tedy speciálně $u \neq f(b)$, takže u je vlastní prefix $f(b)$. Protože $|u| + 1 \geq |f(b)| > |u|$, je $f(b) = ud'$, kde $d' = 1 - d \in B$. $\square$

Tvrzení 60 *Nechť $P \in \Delta(A)$ a nechť pro $a, b \in A$ platí $\forall c \in A \setminus \{a, b\}, P(c) \geq P(a) \geq P(b)$. Uvažujme abecedu $C = A \setminus \{a, b\} \cup \{c\}$, kde $c \notin A$ a sestrojme rozdělení $Q \in \Delta(C)$ předpisem*

$$Q(x) = \begin{cases} P(x) & \textit{pro} \quad x \in A \setminus \{a, b\} \\ P(a) + P(b) & \textit{pro} \quad x = c \end{cases}$$

Nechť $g : C \to B^+$ je minimální prefixový kód pro Q a sestrojme kód $f : A \to B^+$ předpisem

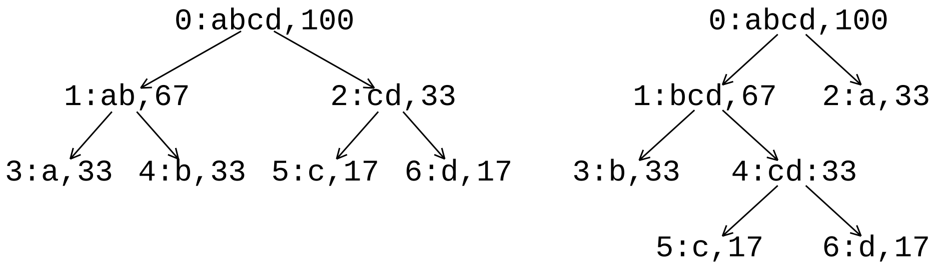
$$f(x) = \begin{cases} g(x) & \textit{pro} \quad x \in A \setminus \{a, b\} \\ g(c)0 & \textit{pro} \quad x = a \\ g(c)1 & \textit{pro} \quad x = b \end{cases}$$

Pak f je minimální prefixový kód pro P.

Důkaz: Předpokládejme sporem že minimální prefixový kód $f' : A \to B^+$ je kratší než f. Vhodnou záměnou písmen lze kód modifikovat tak že jeho délka se nezmění a přitom $f'(a)$, $f'(b)$ jsou kódy s nejmenší délkou, které se liší jen posledním bitem, tj. $f'(a) = u0$, $f'(b) = u1$. Sestrojme kód $g' : C \to B^+$ předpisem $g'(c) = u$ a $g'(x) = f'(x)$ pro $x \in C \setminus \{c\}$. Pak

$$\begin{aligned} L(Q, g') &= \sum_{x \in C} Q(x) \cdot |g'(x)| = \sum_{x \in C \setminus \{c\}} P(x) \cdot |f'(x)| + (P(a) + P(b)) \cdot |u| \\ &\leq L(P, f') - P(a) - P(b) < L(P, f) - P(a) - P(b) = L(Q, g) \end{aligned}$$

a to je spor. $\square$



u	λ	0	1	00	01	10	11
$\ell(u)$	0	1	2	3	4	5	6
$\overline{f}(u)$	$abcd$	ab	cd	a	b	c	d
$\mu(u)$	1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

u	λ	0	1	00	01	010	011
$\ell(u)$	0	1	2	3	4	5	6
$\overline{f}(u)$	$abcd$	bcd	a	b	cd	c	d
$\mu(u)$	1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

Obrázek 15: Ohodnocené binární stromy

Huffmannův algoritmus používá datovou strukturu ohodnoceného binárního stromu. Pro Huffmannovy kódy z obrázku 13 a 14 jsou příslušné ohodnocené stromy na obrázku 15. Ohodnocení vrcholů je zapsáno ve tvaru $\ell(u) : \overline{f}(u), \mu(u)$, kde $\ell(u) = |\{v \in T : v < u\}|$ je pořadí u v lexikografickém uspořádání T, $\overline{f} : T \to \mathcal{P}(A)$ přiřazuje vrcholům podmnožiny abecedy A, a $\mu : T \to [0, 1]$ přiřazuje vrcholům jejich pravděpodobnosti (vyjádřené v procentech).

Definice 22

(1) **Ohodnocený strom** *nad A je trojice $\mathcal{T} = (T, f, \mu)$, kde $T \subseteq B^+$ je úplný binární strom, $\mu : T \to [0, \infty)$ je váhová funkce splňující $\mu(u) = \mu(u0) + \mu(u1)$, a $f : A \to L(T)$ je bijektivní zobrazení.*

(2) Ohodnocený strom $\mathcal{T}$ je **uspořádaný**, *pokud platí $u < v \implies \mu(u) \geq \mu(v)$.*

(3) Spojení ohodnocených stromů $\mathcal{T}_0 = (T_0, f_0, \mu_0)$ a $\mathcal{T}_1 = (T_1, f_1, \mu_1)$ nad disjunktními abecedami A_0, A_1 je ohodnocený strom $\mathcal{T} = (T, f, \mu) = \mathcal{T}_0 \vee \mathcal{T}_1$ nad abecedou $A = A_0 \cup A_1$, kde

(a) $T = 0T_0 \cup 1T_1 \cup \{\lambda\} = \{0u :\ u \in T_0\} \cup \{1u :\ u \in T_1\} \cup \{\lambda\}$,

(b) $\mu(0u) = \mu_0(u)$, $\mu(1u) = \mu_1(u)$, $\mu(\lambda) = \mu_0(\lambda) + \mu_1(\lambda)$

(c) $f(a) = \begin{cases} 0f_0(a) & \textit{pro} \quad a \in A_0 \\ 1f_1(a) & \textit{pro} \quad a \in A_1. \end{cases}$

Ohodnocený strom (T, f, μ) určuje ohodnocení $\overline{f} : T \to \mathcal{P}(A)$ vrcholů stromu podmnožinami abecedy A předpisem

$$\overline{f}(f(a)) = \{a\}, \ \ \overline{f}(u) = \overline{f}(u0) \cup \overline{f}(u1).$$

Definice 23 **Huffmannův algoritmus** *na základě vstupu $P \in \Delta(A)$ vytvoří ohodnocený binární strom následujícím postupem:*

(1) Inicializace: Vytvoř seznamu ohodnocených stromů $(\mathcal{T}_a = (T_a, f_a, \mu_a))_{a\in A}$ nad jednoprvkovými abecedami $\{a\}$. Zde $T_a = \{\lambda\}$, $\mu_a(\lambda) = P(a)$, $f_a(\lambda) = a$.

(2) V seznamu ohodnocených stromů nad disjunktními abecedami (jejichž sjednocení je A) najdi dvojici stromů $\mathcal{T}_0$, $\mathcal{T}_1$ s nejmenšími váhami kořenů $\mu_0(\lambda) \geq \mu_1(\lambda)$ a nahraď je jejich spojením $\mathcal{T}_0 \vee \mathcal{T}_1$.

(3) Opakuj krok (2) dokud nezůstane jediný strom (nad abecedou A)

Konstrukce kódu rozdělení $P = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$ na obrázku 14 probíhá v následujících krocích:

1. $[(\lambda, a, \frac{1}{3})]$, $[(\lambda, b, \frac{1}{3})]$, $[(\lambda, c, \frac{1}{6})]$, $[(\lambda, d, \frac{1}{6})]$
2. $[(\lambda, a, \frac{1}{3})]$, $[(\lambda, b, \frac{1}{3})]$, $[(\lambda, cd, \frac{1}{3}), (0, c, \frac{1}{6}), (1, d, \frac{1}{6})]$
3. $[(\lambda, ab, \frac{2}{3}), (0, a, \frac{1}{3}), (1, b, \frac{1}{3})]$, $[(\lambda, cd, \frac{1}{3}), (0, c, \frac{1}{6}), (1, d, \frac{1}{6})]$
4. $[(\lambda, abcd, 1), (0, ab, \frac{2}{3}), (00, a, \frac{1}{3}), (01, b, \frac{1}{3}), (1, cd, \frac{1}{3}), (10, c, \frac{1}{6}), (11, d, \frac{1}{6})]$

Alternativní konstrukce na obrázku 15 probíhá v následujících krocíchsmallskip

1. $[(\lambda, a, \frac{1}{3})]$, $[(\lambda, b, \frac{1}{3})]$, $[(\lambda, c, \frac{1}{6})]$, $[(\lambda, d, \frac{1}{6})]$
2. $[(\lambda, a, \frac{1}{3})]$, $[(\lambda, b, \frac{1}{3})]$, $[(\lambda, cd, \frac{1}{3}), (0, c, \frac{1}{6}), (1, d, \frac{1}{6})]$
3. $[(\lambda, a, \frac{1}{3})]$, $[(\lambda, bcd, \frac{2}{3}), (0, b, \frac{1}{3}), (1, cd, \frac{1}{3}), (10, c, \frac{1}{6}), (11, d, \frac{1}{6})]$
4. $[(\lambda, abcd, 1), (1, a, \frac{1}{3}), (0, bcd, \frac{2}{3}), (01, b, \frac{1}{3}), (11, cd, \frac{1}{3}), (110, c, \frac{1}{6}), (111, d, \frac{1}{6})]$

Věta 61 *Huffmannův algoritmus z Definice 23 sestrojí uspořádaný strom a jeho kód je minimální.*

Důkaz: Stejně jako v Tvrzení 60 zvolíme písmena $a, b \in A$ s nejmenšími pravděpodobnostmi, tj. taková, že pro každé $c \in A \setminus \{a, b\}$ platí $P(c) \geq P(a) \geq P(b)$. Pro abecedu $C = A \setminus \{a, b\} \cup \{c\}$ a rozdělení Q Huffmannův algoritmus sestrojí ohodnocený strom $\mathcal{T}_C$. Pro ohodnocený strom $\mathcal{T}_A$ rozdělení P platí

$$T_A = T_C \cup \{u0, u1\}, \quad \mu_A(u0) = P(a), \quad \mu_A(u1) = P(b),$$

kde $u = f_C(c)$. Podle indukčního předpokladu T_C je uspořádaný a jeho kód je minimální. Podle Tvrzení 60 je kód T_A také minimální. Ukážeme, že T_A je uspořádaný. Pro dva nové vrcholy $u0, u1$ platí $u0 < u1$ a $\mu_A(u0) = P(a) \geq P(b) = \mu_A(u1)$. Předpokládejme sporem že $\mu(ui) < \mu(vj)$ a přitom $ui < vj$ pro nějaké $i \in B$, $v \in B^* \setminus \{u\}$, $j \in B$. Pak $|u| \leq |v|$ a tedy $u < v$. Protože $\mu(u0) \geq \mu(u1)$, $\mu(v0) \geq \mu(v1)$, dostáváme $\mu(u1) < \mu(v0)$. Protože také $\mu(u0) \leq \mu(v1)$, dostáváme odtud $\mu(u) < \mu(v)$ a to je spor. □

Věta 62 *Je-li $\mathcal{T} = (T, \mu, f)$ uspořádaný strom a $\mu(\lambda) = 1$, je f minimální kód pro rozdělení $P(a) = \mu(f(a))$.*

Důkaz: Nechť $u0, u1 \in T$ jsou dva největší prvky T. Protože T je uspořádaný, mají $u0$ a $u1$ nejmenší pravděpodobnosti a strom $\mathcal{T}'$ který získáme z $\mathcal{T}$ odstraněním $u0, u1$ má rozdělení Q. Podle indukčního předpokladu je $\mathcal{T}'$ minimální kód a podle Tvrzení 14 je $\mathcal{T}$ minimální kód. □

3.4 Kódování spočetných abeced

Teorii kódů lze rozšířit i na spočetné abecedy. Blokový kód spočetné abecedy A je zobrazení $f : A \to B^+$, kde B je binární abeceda. Množina slov nad A je

$$A^* = \bigcup_{k\geq 0} A^k = \{\lambda\} \cup A \cup A^2 \cup \cdots$$

Kód $f : A \to B^*$ lze rozšířit na zobrazení $f^* : A^* \to B^*$. Definice prostého, prefixového a jednoznačného kódu zůstávají v platnosti. I pro nekonečné kódy platí Věty 54, 55 a 57.

Věta 63 *Je-li $f : \mathbb{N} \to B^+$ jednoznačný kód, pak platí Kraftova nerovnost $\sum_{i\in\mathbb{N}} 2^{-|f(i)|} \leq 1$. Naopak je-li $d = (d_i)_{i\in\mathbb{N}}$ posloupnost přirozených čísel, která splňuje Kraftovu nerovnost $\sum_{i\in\mathbb{N}} 2^{-d_i} \leq 1$, pak existuje prefixový kód $f : \mathbb{N} \to B^*$ pro který $|f(i)| = d_i$.*

Důkaz: Nechť $f : \mathbb{N} \to B^+$ je jednoznačný kód. Pak jeho restrikce $f : \mathbb{Z}_n \to B^+$ je také jednoznačný kód a tedy $\sum_{i<n} 2^{-d_i} \leq 1$. Z toho plyne $\sum_{i\in\mathbb{N}} 2^{-d_i} \leq 1$. Naopak předpokládejme že $\sum_{i\in\mathbb{N}} 2^{-d_i} \leq 1$. Pak každá množina $A_k = \{i \in \mathbb{N} : d_i \leq k\}$ je konečná. Nechť k_0 je nejmenší číslo pro které je A_{k_0} neprázdná. Podle důkazu Věty 54 existuje prefixový kód $f : A_{k_0} \to B^+$ s délkami $|f(i)| = d_i$. Je-li již definován prefixový kód $f : A_k \to B^+$, existuje jeho rozšíření na prefixový kód $f : A_{k+1} \to B^+$. Společné rozšíření všech těchto kódů je prefixový kód $f : \mathbb{N} \to B^+$. □

Věta 64 *Nechť $P \in \Delta(\mathbb{N})$ je rozdělení s konečnou entropií. Pak $\mathcal{H}(P) \leq L(P) \leq \mathcal{H}(P) + 1$.*

Tvrzení 65 *Existuje prostý kód $b : \mathbb{N} \to B^*$ pro který $|b(n)| = \lfloor \log(n+1) \rfloor \leq \log(n+1)$ (obrázek 16). Dále existují prefixové kódy $b_k : \mathbb{N} \to B^*$ takové, že*

$$|b_k(n)| = \left\lfloor \frac{n+k}{2^{k-1}} \right\rfloor, \quad k \geq 1$$

n	$b(n)$	$b_1(n)$	$b_2(n)$	$b_3(n)$	$\mathfrak{b}_0(n)$	$\mathfrak{b}_{\mathbb{N}}(n)$	$\mathfrak{l}_1(n)$	$\mathfrak{l}_2(n)$
0	λ	0	00	000	0	0	0	00
1	0	10	01	001	100	1000	1	01
2	1	110	100	010	101	1001		10
3	00	1^30	101	011	11000	10100		11
4	01	1^40	1100	1000	11001	10101		
5	10	1^50	1101	1001	11010	10110		
6	11	1^60	1^300	1010	11011	10111		
7	000	1^70	1^301	1011	1^30000	11000000		
8	001	1^80	1^400	11000	1^30001	11000001		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
14	111	$1^{14}0$	1^700	1^3010	1^30111	11000111		
15	0000	$1^{15}0$	1^701	1^3011	1^400000	110010000		
16	0001	$1^{16}0$	1^800	1^4000	1^400001	110010001		

Obrázek 16: Kódování přirozených čísel

Důkaz: Definujme $b(n)$ jako n-tý prvek v lexikografickém uspořádání binárních slov

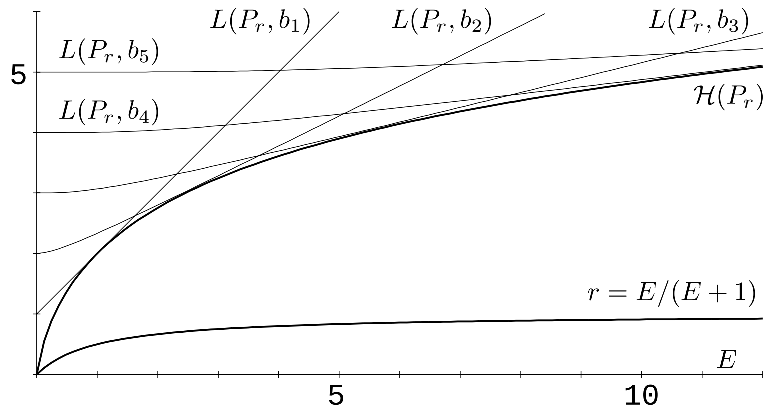
$$B^* = \{\lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, \ldots\}.$$

Pro $b : \mathbb{N} \to B^*$ a inverzní zobrazení $b^{-1} : B^* \to \mathbb{N}$ platí vzorce

$$\begin{aligned} b(n)_i &= \left\lfloor \frac{n - \lfloor \log(n+1) \rfloor}{2^i} \right\rfloor \bmod 2, \quad i < \lfloor \log(n+1) \rfloor \\ b^{-1}(u) &= -1 + 2^{|u|} + 2^{|u|-1}u_0 + 2^{|u|-2}u_1 + \cdots + u_{|u|-1}, \quad u \in B^* \end{aligned}$$

Kód b je prostý ale není ani prefixový ani jednoznačný. Nechť $\mathfrak{l}_k(n)$ je n-tý prvek v lexikografickém uspořádání B^k (obrázek 16). Kódy b_k jsou definovány vzorci $b_1(n) = 1^n0$,

$$b_k(n) = 1^{\lfloor n/2^{k-1} \rfloor} \mathfrak{l}_k(n - \lfloor n/2^{k-1} \rfloor). \ \square$$



Obrázek 17: Kódy pro geometrické rozdělení

Kódy b_k jsou vhodné pro geometrické rozdělení (Příklad 2) $P_r(n) = (1-r)r^n$, kde $0 < r < 1$ se střední hodnotou a entropií

$$\mathbb{E}(P) = \frac{r}{1-r}, \quad \mathcal{H}(P) = \log \frac{1}{1-r} + \frac{r}{1-r} \log \frac{1}{r}.$$

Délky těchto kódů jsou na obrázku 17.

Tvrzení 66 *Pro prefixový kód b_k a geometrické rozdělení P_r je*

$$L(P_r, b_k) = k - 1 + \frac{1}{1 - r^{2^{k-1}}}$$

Důkaz: Položme $j = 2^{k-1}$

$$\begin{aligned} L(P, b_k) &= (1-r)\left(k \cdot (1 + \cdots + r^{j-1}) + (k+1) \cdot r^j \cdot (1 + \cdots + r^{j-1}) + \cdots\right) \\ &= (1 - r^j) \cdot \left(k + (k+1)r^j + (k+1)r^{2j} + \cdots\right) \\ &= k + r^j + r^{2j} + r^{3j} + \cdots = k - 1 + \frac{1}{1-r^j} \quad \square \end{aligned}$$

Speciálně pro $r_k = 2^{-2^{1-k}}$ je $-\log P_{r_k}(n) = -\log(1 - 2^{-2^{1-k}}) + \frac{n}{2^{k-1}}$, $L(P_{r_k}, b_k) = k + 1$. S použitím přibližného vzorce $r_k \approx 1 - 2^{1-k} \cdot \ln(2)$ dostáváme $\mathcal{H}(P_{r_k}) \approx k + 0.972 - 2^{1-k}$, takže délka kódu je velmi blízká entropii.

Ukážeme nyní konstrukci prefixového kódu $\mathfrak{b}_{\mathbb{N}}$, který má stejnou asymptotickou délku jako (nweprefixový) kód b.

Tvrzení 67 *Existují prefixové kódy $\mathfrak{b}_0, \mathfrak{b}_{\mathbb{N}} : \mathbb{N} \to B^+$ takové, že*

$$\begin{aligned} |\mathfrak{b}_0(n)| &\leq 2\log(n+1) + 1 \\ |\mathfrak{b}_{\mathbb{N}}(n)| &\leq \log(n+1) + 2\log(\log(n+1)+1) + 1, \quad \lim_{n\to\infty} \frac{|\mathfrak{b}_{\mathbb{N}}(n)|}{\log(n)} = 1. \end{aligned}$$

Důkaz: Pro kód $\mathfrak{b}_0(n) = b_1(|b(n)|)b(n) = 1^{|b(n)|}0b(n)$ je

$$|\mathfrak{b}_0(n)| = 2|b(n)| + 1 = 2\lfloor \log(n+1) \rfloor + 1 \leq 2\log(n+1) + 1$$

Pro $\mathfrak{b}_{\mathbb{N}}(n) = \mathfrak{b}_0(|b(n)|)b(n)$ pak platí

$$|\mathfrak{b}_{\mathbb{N}}(n)| \leq 2\log(|b(n)|+1) + 1 + |b(n)| \leq \log(n+1) + 2\log(\log(n+1)+1) + 1. \quad \square$$

Obecně pro libovolnou abecedu A lze sestrojit prefixový kód s asymptotickou délkou $|u| \log |A|$.

Tvrzení 68 *Pro každou konečnou abecedu A existuje prefixový kód $\mathfrak{b}_A : A^* \to B^+$ pro který platí*

$$|\mathfrak{b}_A(u)| \leq |u| \cdot \log|A| + 4\log(|u|+1) + 2\log\log|A| + 2, \quad \lim_{|u|\to\infty} \frac{|\mathfrak{b}_A(u)|}{|u|} = \log|A|$$

Speciálně pro binární abecedu platí $|\mathfrak{b}_B(u)| \leq |u| + 4\log(|u|+1) + 2$.

Důkaz: Označíme prvky abecedy A přirozenými čísly, takže $A = \{0, 1, \ldots, m-1\}$. Pro $u \in A^k$ nechť $\ell(u) = u_0 \cdot |A|^{k-1} + \cdots + u_{k-1}$ je pořadí u v lexikografickém uspořádání množiny A^k. Položme $\mathfrak{b}_A(u) = \mathfrak{b}_0(|u|)\mathfrak{b}_{\mathbb{N}}(\ell(u))$. Protože $\ell(u) + 1 \leq |A|^{|u|}$, je

$$\begin{aligned}
|\mathfrak{b}_A(u)| &\leq 2\log(|u|+1) + 1 + \log(|\ell(u)|+1) + 2\log(\log(|\ell(u)|+1)+1) + 1 \\
&\leq 2\log(|u|+1) + |u| \cdot \log|A| + 2\log(|u|\log|A|+1) + 2 \\
&\leq |u| \cdot \log|A| + 2\log(|u|+1) + 2\log((|u|+1)\log|A|) + 2 \\
&\leq |u| \cdot \log|A| + 4\log(|u|+1) + 2\log\log|A| + 2. \ \square
\end{aligned}$$

u	λ	0	1	00	01	10	11	000
$\ell(u)$	0	0	1	0	1	2	3	0
$\mathfrak{b}_B(u)$	00	1000	1001000	1010	1011000	1011001	10110100	110000

Pro $d > 0$ definujeme lexikografický kód $\mathfrak{l}^d : [0, 2^d) \to B^d$. Pro $n < 2^d$ je $\mathfrak{l}_d(n)$ n-tý prvek B^d v lexikografickém uspořádání (obrázek 16).

3.5 Kódy pro stacionární procesy

Definice 24 *Nechť $\mu : A^* \to [0,1]$ je symbolická míra nad abecedou A a $f : A^* \to B^*$ je zobrazení do množiny binárních slov. Horní a dolní délka f při μ je*

$$\underline{L}(\mu, f) = \liminf_{n\to\infty} \frac{1}{n} \sum_{u \in A^n} \mu(u) \cdot |f(u)|, \quad \overline{L}(\mu, f) = \limsup_{n\to\infty} \frac{1}{n} \sum_{u \in A^n} \mu(u) \cdot |f(u)|$$

Příklad 11 *Pro prefixový kód $\mathfrak{b}_A : A^* \to B^+$ a každou symbolickou míru μ platí*

$$\underline{L}(\mu, \mathfrak{b}_A) = \overline{L}(\mu, \mathfrak{b}_A) = \log|A|.$$

Také pro symbolické míry platí Shannonova věta 57, ale v silnějším tvaru: Délka kódu je zdola omezená entropií nejen pro blokové kódy ale pro každé prosté zobrazení $f : A^* \to B^*$ (Věta 69). K entropii se lze délkou blokového kódu libovolně přiblížit (Věta 74). Dokonce existuje prefixový kód $f : A^* \to B^+$, jehož délka je přesně entropie symbolické míry (Věta 75).

Věta 69 *Nechť A je konečná abeceda, $f : A^* \to B^*$ je prosté zobrazení a nechť $\mu : A^* \to [0,1]$ je stacionární míra nad A. Pak $\mathcal{H}(\mu) \leq \underline{L}(\mu, f)$.*

Důkaz: Pro prefixový kód $\mathfrak{b}_B : B^* \to B^*$ sestrojený v Tvrzení 68 platí

$$|\mathfrak{b}_B(u)| \leq |u| + 4\log(|u|+1) + 2.$$

Pro dané $\varepsilon > 0$ existuje n_ε takové že pro všechna $n > n_\varepsilon$ je $|\mathfrak{b}_B(u)| \leq (1+\varepsilon)|u|$. Pro $u \in A^n$ položme $g_n(u) = \mathfrak{b}_B(f(u))$. Pak kód $g_n : A^n \to B^*$ je prefixový. Podle Shannonovy Věty 57 platí

$$\mathcal{H}(\mu_{|A^n}) \leq L(\mu_{|A^n}, g_n) = \sum_{u \in A^n} \mu(u) \cdot |g_n(u)| \leq (1+\varepsilon) \sum_{u \in A^n} \mu(u) \cdot |f(u)|$$

Po vydělení n dostáváme v limitě $\mathcal{H}(\mu) = \lim_{n\to\infty} \mathcal{H}(\mu_{|A^n})/n \leq (1+\varepsilon) \cdot \underline{L}(\mu, f)$ a tato nerovnost platí pro každé $\varepsilon > 0$. $\square$

Věta 70 *Nechť $f : A^* \to B^*$ je prosté zobrazení, $P \in \Delta(A)$ a $(X_n)_{n\geq 0}$ bernoulliovský proces s rozdělením P. Pak*

$$\lim_{n\to\infty} \frac{|f(X_{[0,n)})|}{n} \geq \mathcal{H}(P) \text{ i.p.}$$

Důkaz: Pro každé n platí $|\{u \in A^* : |f(u)| \leq n\}| \leq |\{u \in B^* : |u| \leq n\}| < 2^{n+1}$. Pro dané kladné ε, δ zvolme kladné $\eta < \varepsilon$. Podle Věty 47 o typické množině existuje $n_{\delta,\eta}$ takové, že pro všechna $n > n_{\delta,\eta}$ platí

$$\begin{aligned}
&\mathbb{P}[|f(X_{[0,n)})|/n < \mathcal{H}(P) - \varepsilon] \leq \\
&\quad\leq \mathbb{P}[X_{[0,n)} \notin \mathcal{M}^n_\eta(P)] + \mathbb{P}[X_{[0,n)} \in \mathcal{M}^n_\eta(P) \;\&\; |f(X_{[0,n)})| < n(\mathcal{H}(P) - \varepsilon)] \\
&\quad\leq \delta + 2^{-n\cdot(\mathcal{H}(P)-\eta)} \cdot |\{u \in A^n : |f(u)| < n(\mathcal{H}(P) - \varepsilon)\}| \\
&\quad\leq \delta + 2^{-n\cdot(\mathcal{H}(P)-\eta)} \cdot 2^{n\cdot(\mathcal{H}(P)-\varepsilon)+1} = \delta + 2^{-n(\varepsilon-\eta)+1}
\end{aligned}$$

a tedy $\limsup_{n\to\infty} \mathbb{P}[|f(X_{[0,n)})|/n < \mathcal{H}(P) - \varepsilon] \leq \delta$. Protože to platí pro každé $\delta > 0$, je $\lim_{n\to\infty} \mathbb{P}[|f(X_{[0,n)})|/n > \mathcal{H}(P) - \varepsilon] = 0$. □

Pro konvergenci skoro jistě potřebujeme zesílení Čebyševovy nerovnosti (viz Rényi [25]).

Věta 71 (Bernsteinova nerovnost) *Nechť $(X_i)_{i\geq 0}$ jsou nezávislé, omezené reálné náhodné veličiny s stejnou střední hodnotou $\mathbb{E}(X_i) = E$ a rozptylem $\mathbb{V}(X_i) = D$ pro které existuje $K > 0$ takové že $|X_i - E| \leq K$ pro každé i. Pak pro každé $\varepsilon < D/K$ platí*

$$\mathbb{P}\left[\left|\frac{X_0 + \cdots X_{n-1}}{n} - E\right| \geq \varepsilon\right] \leq 2 \cdot e^{-2n\varepsilon^2/9D}.$$

Z Bernsteinovy nerovnosti dostáváme zesílení odhadu pravděpodobnosti typické množiny

$$\begin{aligned}
\mathcal{M}^n_\varepsilon(P) &= \left\{u \in A^n : \left|\mathcal{H}(P) + \frac{1}{n}\log P^n(u)\right| < \varepsilon\right\} \\
&= \left\{u \in A^n : 2^{-n(\mathcal{H}(P)+\varepsilon)} < P^n(u) < 2^{-n(\mathcal{H}(P)-\varepsilon)}\right\}
\end{aligned}$$

Tvrzení 72 *Nechť $(X_i)_{i\geq 0}$ je bernoulliovský proces s rozdělením $P \in \Delta(A)$ a položme*

$$D = \mathbb{V}(\mathfrak{I}_{X_i}) = -\sum_{a\in A} P(a)\log^2 P(a) - \mathcal{H}(P)^2, \;\; K = \max\{-\log P(a) : a \in A\}$$

Pak pro $\varepsilon < D/K$ platí

$$\mathbb{P}[X_{[0,n)} \in \mathcal{M}^n_\varepsilon(P)] > 1 - 2 \cdot e^{-n\varepsilon^2/9D}.$$

Důkaz: Protože $\mathcal{H}(P) \leq \sum_{a\in A} P(a)\cdot K = K$, je $|\mathfrak{I}_{X_i} - \mathcal{H}(P)| \leq K$. Protože $\mathbb{E}(\mathfrak{I}_{X_i}) = \mathcal{H}(P)$, tvrzení plyne z Bernsteinovy nerovnosti. □

Věta 73 *Nechť $(X_i)_{i\in\mathbb{N}}$ je bernoulliovský proces s rozdělením $P \in \Delta(A)$ a $f : A^* \to B^+$ prosté zobrazení. Pak*

$$\liminf_{n\to\infty} \frac{|f(X_{[0,n)})|}{n} \geq \mathcal{H}(P) \text{ s.j.}$$

Důkaz: Podle Věty 40 stačí ukázat konvergenci suprem v pravděpodobnosti. Pro dané $\varepsilon, \delta > 0$ zvolme kladné $\eta < \varepsilon$ a $n > n_{\delta,\eta}$. Pak

$$\begin{aligned}
&\mathbb{P}[\sup_{m\geq n} |f(X_{[0,m)})|/m < \mathcal{H}(P) - \varepsilon] \leq \\
&\qquad\leq \sum_{m\geq n} \mathbb{P}[|f(X_{[0,m)})|/m < \mathcal{H}(P) - \varepsilon] \\
&\qquad\leq \sum_{m\geq n} \mathbb{P}[X_{[0,m)} \notin \mathcal{M}^m_\eta(P)] + \mathbb{P}[X_{[0,m)} \in \mathcal{M}^m_\eta(P) \ \&\ |f(X_{[0,m)})| < m(\mathcal{H}(P) - \varepsilon)] \\
&\qquad\leq \sum_{m\geq n} 2e^{-n\eta^2/9D} + 2^{-n\cdot(\mathcal{H}(P)-\eta)} \cdot |\{u \in A^n :\ |f(u)| < n(\mathcal{H}(P) - \varepsilon)\}| \\
&\qquad\leq c_0 \cdot c^{-n} + \sum_{m\geq n} 2^{-n\cdot(\mathcal{H}(P)-\eta)} \cdot 2^{n\cdot(\mathcal{H}(P)-\varepsilon)+1} = c_0 \cdot c^{-n} + c_1 \cdot 2^{-n(\varepsilon-\delta)+1}
\end{aligned}$$

kde c, c_0, c_1 jsou kladné konstanty, takže $\lim_{n\to\infty} \mathbb{P}[\sup_{m\geq n} |f(X_{[0,m)})|/m > \mathcal{H}(P) - \varepsilon] = 0$. Podle Věty 40 je

$$\liminf_{n\to\infty} \frac{|f(X_{[0,n)})|}{n} \geq \mathcal{H}(P) \text{ s.j.} \quad \square$$

Věta 74 *Pro každou stacionární míru μ nad A a pro každé $\varepsilon > 0$ existuje $n \in \mathbb{N}$ a blokový kód $f_n : A^n \to B^+$ takový že*

$$\frac{L(\mu_{|A^n}, f_n)}{n} \leq \mathcal{H}(\mu) + \varepsilon$$

Důkaz: Existuje n takové že $\frac{1}{n} < \frac{\varepsilon}{2}$ a $\mathcal{H}(\mu_{|A^n})/n < \mathcal{H}(\mu) + \frac{\varepsilon}{2}$. Pro $\mu_{|A^n}$ existuje prefixový kód $f_n : A^n \to B^+$ pro který $L(\mu_{|A^n}, f_n) \leq \mathcal{H}(\mu_{|A^n}) + 1$, takže

$$\frac{L(\mu_{|A^n}, f_n)}{n} < \frac{\mathcal{H}(\mu_{|A^n}) + 1}{n} < \mathcal{H}(\mu) + \varepsilon. \quad \square$$

Věta 75 *Pro každou stacionární míru μ nad A existuje prefixový kód $\mathfrak{b}_\mu : A^* \to B^*$, takový že $\mathcal{H}(\mu) = \underline{L}(\mu, \mathfrak{b}_\mu) = \overline{L}(\mu, \mathfrak{b}_\mu)$.*

Důkaz: Podle Věty 57 pro každé n existuje blokový prefixový kód $f_n : A^n \to B^*$ takový že $\sum_{u\in A^n} \mu(u) \cdot |f_n(u)| \leq \mathcal{H}(\mu_{|A^n}) + 1$. Pro $u \in A^n$ položme $\mathfrak{b}_\mu(u) = \mathfrak{b}_0(n) f_n(u)$, kde $\mathfrak{b}_0$ je prefixový kód z Tvrzení 67. Pak $\mathfrak{b}_\mu : A^* \to B^*$ je prefixový kód a platí

$$\begin{aligned}
\sum_{u\in A^n} \mu(u) \cdot |\mathfrak{b}_\mu(u)| &\leq \sum_{u\in A^n} \mu(u) \cdot (|f_n(u)| + 2\log(n+1) + 1) \\
&\leq \mathcal{H}(\mu_{|A^n}) + 2\log(n+1) + 2 \\
\limsup_{n\to\infty} \frac{1}{n} \sum_{u\in A^n} |\mathfrak{b}_\mu(u)| \cdot \mu(u) &\leq \mathcal{H}(\mu)
\end{aligned}$$

takže $\overline{L}(\mu, f) \leq \mathcal{H}(\mu)$. Podle Věty 69 je $\mathcal{H}(\mu) \leq \underline{L}(\mu, f)$. $\square$

4 Univerzální kódy

4.1 Frekvenční kód

Optimální kódy, které jsme dosud sestrojovali, byly vždy určeny pro dané pravděpodobnostní rozdělení - buď pro $P \in \Delta(A)$, nebo pro symbolickou míru μ. Existují však také "univerzální" kódy, které dobře komprimují celé třídy stacionárních procesů, například všechny bernoulliovské procesy. Nejjednodušší kód tohoto typu je frekvenční kód, který popisuje dané slovo tak, že udá počty výskytů jednotlivých písmen, spolu s pořadím daného slova v množině všech slov se stejnými počty výskytů. Připomeňme, že frekvenční rozdělení slova $u \in A^*$ je $\mathfrak{P}_u(a) = |u|_a/|u|$. Uvažujme lineární uspořádání na A a jemu příslušné lexikografické uspořádání na A^*.

$$u < v \iff |u| < |v| \text{ nebo } |u| = |v| \text{ a } \exists i < |u|, u_{[0,i)} = v_{[0,i)}, u_i < v_i$$

Pro binární abecedu je $\lambda < 0 < 1 < 00 < 01 < 10 < 11 < 000 < \cdots$

Tvrzení 76 *Pro každou abecedu A existuje frekvenční prefixový kód $\mathfrak{f}_A : A^* \to B^*$ takový že pro všechna $u \in A^*$ platí*

$$|\mathfrak{f}_A(u)| \leq |u| \cdot \mathcal{H}(\mathfrak{P}_u) + 2(|A|+1)\log(|u|+1) + |A| + \log\log|A|.$$

Důkaz: Předpokládejme že abeceda A je uspořádaná a uvažujme lexikografické uspořádání na A^*. Známe-li počty písmen $|u|_a$ ve slově $u \in A^*$, je slovo u jednoznačně určeno svým pořadím

$$\ell(u) = |\{v \in A^{|u|} : \; v < u \;\&\; \forall a \in A, |v|_a = |u|_a\}|$$

v množině slov se stejnými frekvencemi jako u. Platí $\ell(u) < |\mathcal{C}(u)| \leq 2^{|u|\cdot\mathcal{H}(\mathfrak{P}_u)} \leq |A|^{|u|}$. Položme

$$\mathfrak{f}_A(u) = \mathfrak{b}_0(|u|_0)\ldots\mathfrak{b}_0(|u|_{|A|-1})\mathfrak{b}_{\mathbb{N}}(\ell(u))$$

Zřejmě f_A je prefixový kód a platí

$$\begin{aligned}
|\mathfrak{f}_A(u)| &\leq |A|\cdot(2\log(|u|+1)+1) + \log(|\ell(u)|+1) + 2\log(\log(|\ell(u)|+1)+1) + 1\\
&\leq |A|\cdot(2\log(|u|+1)+1) + |u|\cdot\mathcal{H}(\mathfrak{P}_u) + 2\log(|u|\log|A|+1) + 1\\
&\leq |u|\cdot\mathcal{H}(\mathfrak{P}_u) + |A|(2\log(|u|+1)+1) + 2\log((|u|+1)\cdot\log|A|)\\
&\leq |u|\cdot\mathcal{H}(\mathfrak{P}_u) + |A|(2\log(|u|+1)+1) + 2\log((|u|+1)\cdot\log|A|)\\
&\leq |u|\cdot\mathcal{H}(\mathfrak{P}_u) + 2(|A|+1)\log(|u|+1) + |A| + \log\log|A|.
\end{aligned}$$

Věta 77 *Nechť $\mathfrak{f}_A : A^* \to B^*$ je frekvenční kód z Tvrzení 76. Pak pro každý bernoulliovský proces X platí*

$$\lim_{n\to\infty} \frac{1}{n}|\mathfrak{f}_A(X_{[0,n)})| = \mathcal{H}(X) \textsf{ s.j.}$$

Důkaz: Podle Tvrzení 76 platí $\limsup_{n\to\infty} \frac{|\mathfrak{f}_A(X_{[0,n)})|}{n} \leq \limsup_{n\to\infty} \mathcal{H}(\mathfrak{P}_{X_{[0,n)}})$. Podle Věty 43 platí $\lim_{n\to\infty} \mathcal{H}(\mathfrak{P}_{X_{[0,n)}}) = \mathcal{H}(P)$ **s.j.** takže

$$\limsup_{n\to\infty} \frac{|f_A(X_{[0,n)})|}{n} \leq \mathcal{H}(P) \textsf{ s.j.}$$

Opačnou nerovnost dává Věta 73. □

Frekvenční kód je univerzální v tom smyslu, že dosahuje optimální kompresi dat pro každý bernoulliovský proces. Na rozdíl od blokových kódů má však jednu velkou nevýhodu. Kódovací i dekódovací algoritmus má exponenciální výpočtovou složitost.

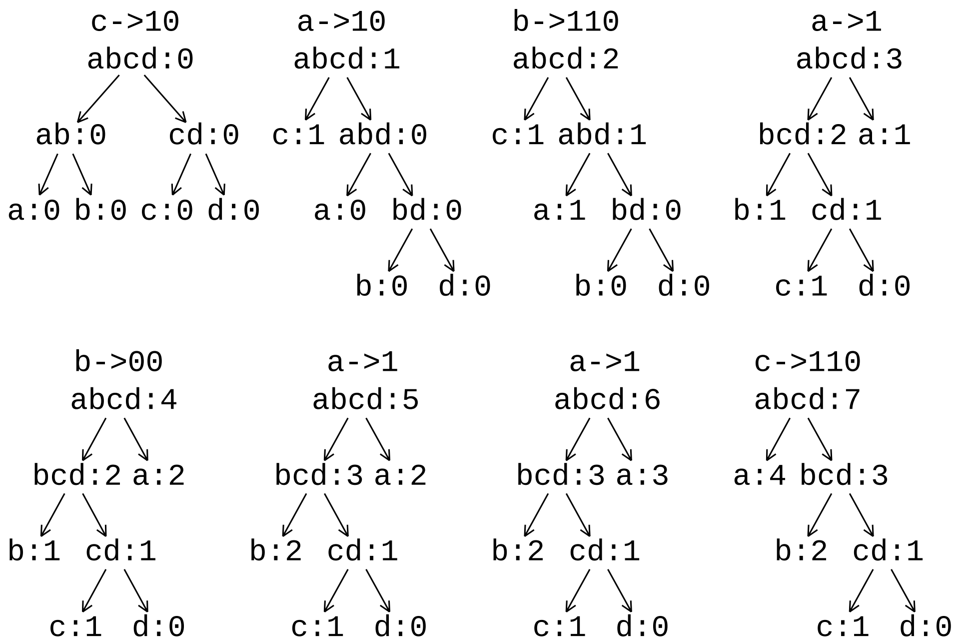
4.2 Adaptivní Huffmannův kód

Další možností jak efektivně kódovat zprávu $u \in A^*$ s libovolným rozdělením je nejprve určit její rozdělení $\mathfrak{P}_u$, sestrojit pro $\mathfrak{P}_u$ minimální kód, odeslat nejprve (v nějakém standardním kódování) tento kód a potom kódovanou zprávu. Existuje ale efektivnější algoritmus, který je pouze jednoprůchodový. Tento algoritmus mění minimální kód dynamicky podle počtů písmen v dosud přečtené zprávě. Dekódovací algoritmus mění minimální kód přesně stejným způsobem, takže je schopen zprávu správně dekódovat. Pro implementaci této metody je třeba odstranit nejednoznačnost konstrukce Huffmannova kódu, která nastává při stejných pravděpodobnostech.

Uvažujme lineárně uspořádanou abecedu A. Toto uspořádání určuje lexikografické uspořádání na A^*. Zavedeme ještě uspořádání množiny $\mathcal{P}(A)$ všech podmnožin A. Pro množinu $x \subseteq A$ označme $\overline{x} \in A^*$ konkatenaci prvků množiny x ve vzestupném pořadí. Toto vnoření přenáší lexikografické uspořádání na $\mathcal{P}(A)$: $x < y$ pokud $\overline{x} < \overline{y}$.

Definice 25 *Uspořádaný ohodnocený strom $\mathcal{T} = (T, f, w)$ je A uspořádaný, pokud pro každé $u < v \in T$ platí $\overline{f}(u) < \overline{f}(v)$.*

Huffmannův algortimus z Definice 23 modifikujeme tak, že v případě $\mu_0(\lambda) = \mu_1(\lambda)$ požadujeme $\overline{f}_i(\lambda) < \overline{f}_0(\lambda) < \overline{f}_1(\lambda)$ pro $i > 1$. Tímto způsobem je Huffmannův strom daného rozdělení P na uspořádané abecedě A určen jednoznačně. Dynamický Huffmannův algoritmus používá místo reálných vah (pravděpodobností) celočíselné váhy, které mají význam počtů písmen v dosud přečteném úseku vstupního slova.



Obrázek 18: Adaptivní Huffmannův kód

Příklad konstrukce kódu je na obrázku 18. Algoritmus převádí vstupní slovo *cababaac* na $10, 10, 110, 1, 00, 1, 1, 110$ (čárky jsou doplněny pouze pro přehlednost). Na začátku jsou všechny váhy nulové, a všechny kódy jsou stejně dlouhé. První písmeno c je kódováno slovem 10. Po jeho přečtení se upraví váha tohoto vrcholu na $\mu(10) = 1$. Aby zůstala zachována vlastnost aditivity vah, je třeba upravit $\mu(1) = 1$ a $\mu(\lambda) = 1$. Po tomto upravení zkontrolujeme, zda je strom uspořádaný, tj. zda pro $u > v$ platí $\mu(u) < \mu(v)$. V našem případě tato podmínka neplatí, takže je třeba vytvořit nový Huffmannův kód. Další vstupní písmeno a má pak kód 10, a po úpravě vah dostaneme $\mu(10) = 1$, $\mu(1) = 1$ a $\mu(\lambda) = \mu(0) + \mu(1) = 2$ (ostatní váhy jsou nulové). Ověříme, že tento strom je uspořádaný, takže můžeme pokračovat dalším písmenem b, které má kód 110. Po tomto kroku je opět třeba přepočítat Huffmannův kód.

Definice 26 Adaptivní Huffmannův algoritmus *uspořádané abecedy A vytvoří na základě vstupu $u \in A^*$ kód $\mathfrak{h}(u) \in B^*$.*

(1) Inicializace: Vytvoř jediný A-uspořádaný strom s váhami $\mu(u) = 0$.

(2) Pro daný A-uspořádaný strom $\mathcal{T} = (T, \mu, f)$ a vstupní písmeno $a \in A$ vypiš na výstup jeho kód $f(a)$.

(3) Zvětši váhu listu $f(a)$ a všech jeho prefixů o jednotku.

(4) Ověř zda upravený strom je uspořádaný, tj. zda splňuje $u > v \implies \mu(u) \leq \mu(v)$. Pokud je podmínka splněna přejdi na (2)

(5) Pokud podmínka splněna není, vytvoř jediný A-uspořádaný strom s váhami listů $\mu(f(a))$ a přejdi na (2)

Huffmannův adaptivní algoritmus určuje prosté zobrazení $\mathfrak{h}_A : A^* \to B^*$. Vyhradíme-li v abecedě A jeden speciální znak "eof", který použijeme na ukončení zprávy, stane se z $\mathfrak{h}_A$ prefixový kód.

Věta 78 *Pro bernoulliovský proces $(X_i)_{i \geq 0}$ s rozdělením $P \in \Delta(A)$, adaptivní Huffmannův kód skoro jistě konverguje k minimálnímu kódu rozdělení P. Délka $|\mathfrak{h}_A(X_{[0,n)})|/n$ skoro jistě konverguje k $\mathcal{H}(P)$.*

Důkaz plyne z Věty 43 podle které frekvence $\mathfrak{P}_{X_{[0,n)}}$ konvergují skoro jistě k P. Při implementaci Huffmannova adaptivního algoritmu lze zvýšit výpočetní rychlost tak, že se podmínka uspořádanosti nekontroluje v každém kroku, takže se přepočítávání minimálního kód provádí méně často.

4.3 Ziv-Lempelův rekurenční kód

Další typ kódů je založen na rekurenci. Podslovo kódovaného textu je určeno adresou svého předcházejícího výskytu. Tyto kódy jsou stejně jako frekvenční a adaptivní Huffmannův kód univerzální: Dosahují optimální kompresi dat pro každý bernoulliovský proces. Časová složitost kódovacího algoritmu je $cn \log n$, kde c je konstanta a n je délka vstupního slova. Časová složitost dekódovacího algoritmu je dokonce lineární. Rekurenční kódy lze konstruovat pro libovolnou abecedu. Ukážeme si konstrukci rekurenčního kódu pro binární abecedu, pro kterou je důkaz optimality jednodušší. Dané slovo $x \in B^*$ rozložíme na bloky proměnné délky tak, že začneme prázdným slovem a pokračujme vždy nejkratším úsekem, který se mezi předcházejícími bloky nevyskytuje. Této reprezentaci říkáme **rozbor**. Například rozbor slova $x = 000011001110011001101100$ je $R(x) = y = \lambda, 0, 00, 01, 1, 001, 11, 0011, 00110, 111, 00$.

i	0	1	2	3	4	5	6	7	8	9	10
$y^{(i)}$	λ	0	00	01	1	001	11	0011	00110	111	00
a_i		0	1	1	0	2	4	5	7	6	1
c_i		0	0	1	1	1	1	1	0	1	0
k_i	0	1	3	5	6	9	11	15	20	23	25

Definice 27 *Rozbor slova $x \in B^*$ je posloupnost binárních slov $y = (y^{(i)})_{i \le C(x)}$, kde $y^{(0)} = \lambda$, a $y^{(i)} = x_{[k_{i-1}, k_i)}$ pro $i > 0$. Zde $k_0 = 0$ a*

$$k_{i+1} = \min\{j \in (k_i, |u|] : \forall m \le i, x_{[k_i,j)} \ne y^{(m)}\}.$$

Je-li množina $\{j \in (k_i, |u|] : \forall m \le i, x_{[k_i,j)} \ne y^{(m)}\}$ prázdná, je $k_{i+1} = |u|$ a $C(x) = i$.

Označme c_i poslední bit slova $y^{(i)}$. Jeho prefix délky $k_i - k_{i-1} - 1$ je nějaké $y^{(a_i)}$, kde $a_i < i$, takže $y^{(i)} = y^{(a_i)} c_i$. Kódujeme-li každé a_i jako binární slovo délky $d = \lceil \log C(x) \rceil$, dostáváme rekurenční kód

$$\mathfrak{r}(x) = \mathfrak{l}_d(a_1)c_1 \mathfrak{l}_d(a_2)c_2 \ldots \mathfrak{l}_d(a_{C(x)})c_{C(x)}$$

kde $\mathfrak{l}_d(n)$ je n-té slovo abecedy B^d v lexikografickém uspořádání (viz obrázek 16). Například pro uvažované slovo x dostáváme $C(x) = 10$ a $d = 4$, takže

$$\begin{aligned}
x &= 0, 00, 01, 1, 001, 11, 0011, 00110, 111, 00 \\
(a_i, c_i)_{i \le 10} &= (0,0), (1,0), (1,1), (0,1), (2,1), (4,1), (5,1), (7,0), (6,1), (1,0) \\
\mathfrak{r}(x) &= \overbrace{0000}^{0} 0 \overbrace{0001}^{1} 0 \overbrace{0001}^{1} 1 \overbrace{0000}^{0} 1 \overbrace{0010}^{2} 1 \overbrace{0100}^{4} 1 \overbrace{0101}^{5} 1 \overbrace{0111}^{7} 0 \overbrace{0110}^{6} 1 \overbrace{0001}^{1} 0
\end{aligned}$$

Všimněme si, že rekurenční kód je nejefektivnější pro **Champernownovu sekvenci**, která vznikne jako konkatenace všech slov B^+ v lexikografickém pořadí:

$$\begin{aligned}
R(x) &= \lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, \ldots \\
(a_i, c_i) &= (0,0), (0,1), (1,0), (1,1), (2,0), (2,1), (3,0), (3,1), (4,0), \ldots
\end{aligned}$$

Pro délku rekurenčního kódu slova $x \in B^*$ platí

$$|\mathfrak{r}(x)| = C(x)(\lceil \log(C(x) + 1) \rceil + 1) < C(x) \log(C(x) + 3)$$

Ukážeme, že pro každý bernoulliovský proces X s rozdělením $P \in \Delta(B)$, $|\mathfrak{r}(X_{[0,n)})|/n$ konverguje skoro jistě k entropii $\mathcal{H}(X) = \mathcal{H}(P)$.

Lemma 79 *Nechť $x \in B^*$. Za předpokladu, že jmenovatel zlomku je kladný platí*

$$C(x) < \frac{|x|}{\log|x| - \log\log|x| - 3} \implies \lim_{|x| \to \infty} \frac{C(x)}{|x|} = 0$$

Důkaz: Předpokládejme nejprve že $|x| = n_k = \sum_{j=1}^{k} j \cdot 2^j = (k-1)2^{k+1} + 2$. Pak $C(x)$ nabývá svou největší hodnotu pokud se v rozboru x vyskytují všechna slova délky nejvýše k (Champernownova sekvence), takže

$$|x| = n_k \implies C(x) \le \sum_{j=1}^{k} 2^j < 2^{k+1} < \frac{n_k}{k-1}$$

Je-li $n_k \le |x| < n_{k+1}$, je $2^{k+1} < n_k \le |x|$, takže $k + 1 \le \log|x|$ a

$$|x| < k2^{k+2} + 2 < (k+1)2^{k+2} \le \log|x| \cdot 2^{k+2}$$

takže $k + 2 > \log|x| - \log\log|x|$. Délka rozboru $C(x)$ nabývá svou největší hodnotu pokud se v rozboru vyskytují všechna slova délky nejvýše k a žádné slovo délky $k+2$. Platí tedy

$$C(x) \le \frac{n_k}{k-1} + \frac{|x| - n_k}{k+1} \le \frac{|x|}{k-1} \le \frac{|x|}{\log|x| - \log\log|x| - 3}. \quad \square$$

Pro $l > 0$ položme

$$C_x(l) = \{i \leq C(x) :\ |y^{(i)}| = l\},\ \ Q_x(l) = |C_x(l)|/C(x)$$

Pak $Q_x \in \Delta(\mathbb{N})$ je rozdělení na kladných přirozených číslech, které má jen konečně mnoho nenulových hodnot.

Lemma 80 *Nechť* $P \in \Delta(B)$. *Pak*

$$\log P(x) \leq -\sum_{l=1}^{\infty} |C_x(l)| \cdot \log |C_x(l)|.$$

Důkaz: Protože log je konkávní funkce, platí $\sum_{i=0}^{n-1} q_i \log x_i \leq \log(\sum_{i=0}^{n-1} q_i x_i)$ kdykoliv q je pravděpodobnostní vektor. Pro rozbor y platí

$$\begin{aligned}\log P(x) &= \sum_{i=1}^{C(x)} \log P(y^{(i)}) = \sum_{l=1}^{\infty} |C_x(l)| \sum_{i \in C_x(l)} \frac{1}{|C_x(l)|} \log P(y^{(i)}) \\ &\leq \sum_{l=1}^{\infty} |C_x(l)| \log \left(\sum_{i \in C_x(l)} \frac{P(y^{(i)})}{|C_x(l)|} \right) \leq \sum_{l=1}^{\infty} |C_x(l)| \cdot \log \frac{1}{|C_x(l)|}\end{aligned}$$

Pokud totiž $i, j \in C_x(l)$, pak buď $i = j$ nebo $y^{(i)} \neq y^{(j)}$ takže $\sum_{i \in C_x(l)} P(y^{(i)}) \leq 1$. $\square$

Lemma 81

$$-\frac{\log P(x)}{|x|} \geq \frac{C(x) \log C(x)}{|x|} + \frac{C(x)}{|x|} \log \frac{C(x)}{|x|} - \left(1 + \frac{C(x)}{|x|}\right) \log \left(1 + \frac{C(x)}{|x|}\right)$$

Důkaz: Střední hodnota rozdělení Q_x je $E = \sum_{i \geq 1} l \cdot |C_x(l)|/C(x) = |x|/C(x)$. Podle Věty 10 dostáváme

$$\log C(x) - \sum_{l=1}^{\infty} \frac{|C_x(l)|}{C(x)} \log |C_x(l)| = \mathcal{H}(Q_x) \leq (E+1) \log \frac{E+1}{E} + \log E$$

Po vynásobení $C(x)/|x|$ dostáváme

$$\frac{C(x) \cdot \log C(x)}{|x|} - \frac{C(x)}{|x|} \sum_{l=1}^{\infty} \frac{|C_x(l)|}{C(x)} \cdot \log |C_x(l)| \leq \frac{E+1}{E} \log \frac{E+1}{E} + \frac{\log E}{E}$$

Podle Lemma 80 je

$$\begin{aligned}-\frac{\log P(x)}{|x|} &\geq \frac{C(x)}{|x|} \sum_{l=1}^{\infty} \frac{C(x)_l}{C(x)} \log C(x)_l \\ &\geq \frac{C(x) \log C(x)}{|x|} - \frac{E+1}{E} \log \frac{E+1}{E} + \frac{1}{E} \log \frac{1}{E}\end{aligned}$$

Po dosazení $E = |x|/C(x)$ dostaneme požadovanou nerovnost. $\square$

Věta 82 *Nechť* $X = (X_i)_{i \geq 0}$ *je bernoulliovský proces s rozdělením* $P \in \Delta(B)$. *Pak*

$$\lim_{n \to \infty} \frac{|\mathfrak{r}(X_{[0,n)})|}{n} = \mathcal{H}(X) \text{ s.j.}$$

Důkaz: Protože $C(x)/|x| = 1/E$ se blíží k nule pro $|x|$ jdoucí k nekonečnu, plyne z Lemma 81

$$\limsup_{|x|\to\infty}\left(\frac{C(x)\log C(x)}{|x|}+\frac{\log P(x)}{|x|}\right)\leq 0,\quad \limsup_{|x|\to\infty}\frac{|\mathfrak{r}(x)|+\log P(x)}{|x|}\leq 0$$

Protože $-\log P(X_{[0,n)})/n$ konverguje skoro jistě k $\mathcal{H}(X)$, dostáváme

$$\limsup_{n\to\infty}\frac{|\mathfrak{r}(X_{[0,n)})|}{n}\leq \mathcal{H}(X)\ \textsf{s.j.}$$

Protože $\mathfrak{r}: B^* \to B^*$ je prosté zobrazení, plyne z Věty 73

$$\lim_{n\to\infty}\frac{|\mathfrak{r}(X_{[0,n)})|}{n} = \mathcal{H}(X)\ \textsf{s.j.}\quad \square$$

Kódovací algoritmus kódu $\mathfrak{r}$ potřebuje dva průchody vstupního slova. Po prvním průchodu určí délku kódu $d = \lceil C(x)\rceil$ a při druhém konstruuje vlastní kód. Kódujeme-li doby návratu pomocí prefixového kódu $\mathfrak{b}_{\mathbb{N}}$, stačí pouze jeden průchod. Přitom je efektivnější kódovat relativní adresy, tj. $i - a_i$ místo a_i. Protože $i - a_i > 0$, lze nulu využít pro označení konce kódu. Dostáváme tak prefixový kód

$$\mathfrak{r}'(x) = \mathfrak{b}_{\mathbb{N}}(1-a_1)c_1\mathfrak{b}_{\mathbb{N}}(2-a_2)\cdots\mathfrak{b}_{\mathbb{N}}(C(x)-a_{C(x)})c_{C(x)}\mathfrak{b}_{\mathbb{N}}(0).$$

Rekurenční kód je optimální nejen pro každý bernoulliovský proces, ale obecněji pro každý stacionární **ergodický** proces. Symbolická míra se nazývá ergodická, pokud není vlastní konvexní kombinací žádných dvou jiných symbolických měr. Například markovský proces s přechodovou maticí R je ergodický právě když R je **ireducibilní** matice, to znamená, že pro každé $a, b \in A$ existuje $n > 0$ takové, že $R^n(a,b) > 0$. (Každá primitivní matice je ireducibilní, naopak to však neplatí.). Pro každý ergodický stacionární proces nyní platí, že $|\mathfrak{r}(X_{[0,n)})|/n$ konverguje skoro jistě k $\mathcal{H}(X)$.

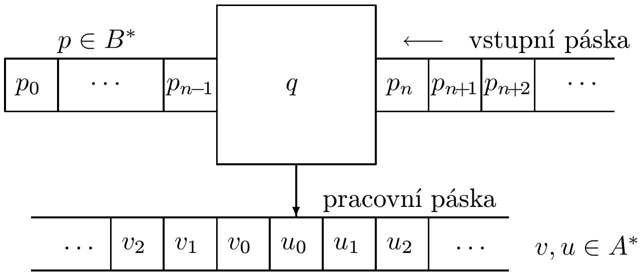
4.4 Algoritmická složitost

Algoritmická složitost slova je délka nejkratšího programu, který toto slovo vypočítá. Slovo je algoritmicky složité, je-li samo svým nejkratším popisem. Slova dané abecedy A tedy kódujeme pomocí programů, což jsou slova v abecedě B daného programovacího jazyka. Ne každé slovo abecedy B ale určuje slovo abecedy A, protože výpočet se může zacyklit. Kódování slov pomocí programů je částečně rekursivní funkce $M : B^* \to A^*$, jejíž definiční obor $\mathcal{D}(M)$ je algoritmicky nerozhodnutelná podmnožina B^*. To znamená, že neexistuje algoritmus, který by rozhodoval zda dané slovo náleží do $\mathcal{D}(M)$. Podstatným principem teorie algoritmické složitosti je to, že program určuje sám svůj konec, říkáme že je sebeomezující. Žádné prodloužení korektního programu (tj, slova v $\mathcal{D}(M)$) není korektní program. To znamená že množina $\mathcal{D}(M)$ je (nekonečný) prefixový kód a lze na ní použít Kraftovu nerovnost. Sebeomezující podmínku lze zaručit tak, že uvažujeme Turingův automat se vstupní páskou, po které se automat nemůže vracet (Obrázek 19).

Pro abecedu A označme $\overline{A} = A \cup \{\lambda\}$ její rozšíření o prázdné slovo. i-té písmeno slova $u \in A^*$ značíme u_i. Je-li $i > |u|$, klademe $u_i = \lambda$. Zobrazení $\sigma : A^* \to A^*$ je definováno předpisem $\sigma(u)_i = u_{i+1}$, takže $|\sigma(u)| = |\sigma(u)| - 1$ pro $|u| > 0$ a $\sigma(\lambda) = \lambda$. Označme $B = \{0, 1\}$ binární abecedu ve které jsou zapsány programy na vstupní pásce, a abecedu $A \supseteq B$ kterou používá pracovní páska.

Definice 28 **Turingův automat** *M nad abecedou A je dán konečnou množinou Q vnitřních stavů, počátečním stavem $q_0 \in Q$ a částečnou přechodovou funkcí*

$$\delta : Q \times \overline{A} \times \overline{B} \to \overline{Q} \times A \times \{-1, 0, 1\}$$



Obrázek 19: Turingův automat

s definičním oborem $\mathcal{D}(\delta) \subset Q \times \overline{A} \times \overline{B}$ *takovou, že pro každé* $(q, a) \in Q \times \overline{A}$ *platí*

$$(q, a, \lambda) \in \mathcal{D}(\delta) \implies (q, a, 0) \notin \mathcal{D}(\delta) \text{ a } (q, a, 1) \notin \mathcal{D}(\delta)$$

Pro daný vnitřní stav $q \in Q$ a písmeno $a \in \overline{A}$ pracovní pásky automat buď přečte písmeno ze vstupní pásky, takže $(q, a, 0)$ nebo $(q, a, 1)$ náleží do $\mathcal{D}(\delta)$ a (q, a, λ) nikoliv, nebo automat žádné vstupní písmeno nečte, takže $(q, a, \lambda) \in \mathcal{D}(\delta)$. Automat se zastaví pokud dojde do nulového stavu $\lambda \in \overline{Q} \setminus Q$, nebo pokud nemůže pokračovat, tj. není-li žádné pravidlo použitelné. **Konfigurace** automatu je trojice (q, v, u), kde $q \in \overline{Q}$ je vnitřní stav a $v, u \in A^*$ je obsah pracovní pásky nalevo a napravo od čtecí hlavy. Čtené písmeno pracovní pásky je u_0, což může být také λ pokud $u = \lambda$. Přechodová funkce určuje nekonečný označený graf, jehož vrcholy jsou konfigurace a hrany jsou označeny prvky $\overline{A}$. Je-li (q, v, u) konfigurace a $\delta(q, u_0, b) = (q', a, z)$, pak existuje označená hrana

$$\begin{array}{lll} (q, v, u) \xrightarrow{b} (q', \sigma(v), v_0 a \sigma(u)) & \text{pokud} & z = -1 \\ (q, v, u) \xrightarrow{b} (q', v, a\sigma(u)) & \text{pokud} & z = 0 \\ (q, v, u) \xrightarrow{b} (q', av, \sigma(u)) & \text{pokud} & z = 1 \end{array}$$

Je-li tedy $(q, u_0, \lambda) \in \mathcal{D}(\delta)$, vede z konfigurace (q, v, u) jediná hrana označená λ, v opačném případě z ní vedou nejvýše dvě hrany označené 0 nebo 1. Každá cesta v konfiguračním grafu je označená slovem $p \in B^*$, které je konkatenací označení jednotlivých hran cesty.

Pro každé $u \in A^*$ definujme částečnou funkci $M_u : B^* \to A^*$ předpisem

$$M_u(p) = x \iff \exists v \in A^*, (q_0, \lambda, u) \xrightarrow{p} (\lambda, v, x).$$

To znamená že $M_u(p) = x$ právě když existuje cesta z počáteční konfigurace (q_0, λ, u) do terminální konfigurace, ve které je stav hlavy $\lambda \in \overline{Q}$ a na pracovní pásce napravo od čtecí hlavy je slovo x. Funkce M_u je částečně rekursivní a její definiční obor je (nekonečný) prefixový kód. Je-li $p \in \mathcal{D}(M_u)$ a $q \sqsubseteq_p p$ (prefix), pak $q \notin \mathcal{D}(M_u)$.

Příklad 12 *Existuje Turingův automat pro který* $M_\lambda(b_1(n)) = b(n)^{-1}$

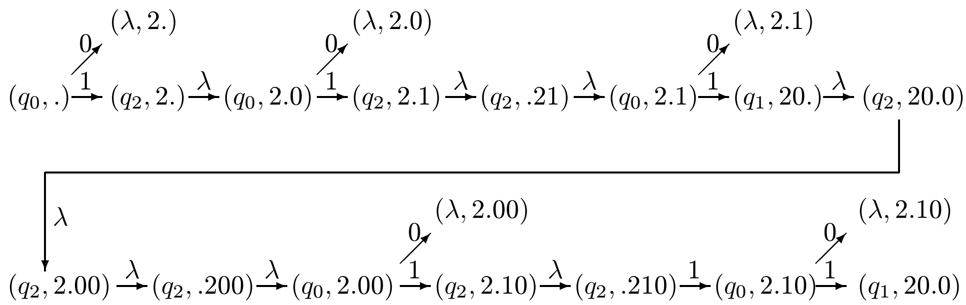
n	0	1	2	3	4	5	6	7	8
$b_1(n)$	0	10	110	1110	11110	$1^5 0$	$1^6 0$	$1^7 0$	$1^8 0$
$b(n)^{-1}$	λ	0	1	00	10	01	11	000	100

Zde $b_1(n) = 1^n 0$ je prefixový kód čísla n a $b(n)$ je inversní slovo k n-tému slovu v lexikografickém uspořádání. Stavová množina automatu je $Q = \{q_0, q_1, q_2\}$ a abeceda je

$A = \{0, 1, 2\}$. Písmeno 2 je pomocné a označuje začátek binárního slova. Přechodová funkce je

$$
\begin{array}{llll}
\delta(q_0, \lambda, 0) : (\lambda, 2, 1) & \delta(q_0, 0, 0) : (\lambda, 0, 0) & \delta(q_0, 1, 0) : (\lambda, 1, 0) & \\
\delta(q_0, \lambda, 1) : (q_2, 2, 1) & \delta(q_0, 0, 1) : (q_2, 1, 0) & \delta(q_0, 1, 1) : (q_1, 0, 1) & \\
\delta(q_1, \lambda, \lambda) : (q_2, 0, 0) & \delta(q_1, 0, \lambda) : (q_2, 1, 0) & \delta(q_1, 1, \lambda) : (q_1, 0, 1) & \delta(q_1, 2, \lambda) : (q_0, 0, 0) \\
\delta(q_2, \lambda, \lambda) : (q_0, 0, 0) & \delta(q_2, 0, \lambda) : (q_2, 0, -1) & \delta(q_2, 1, \lambda) : (q_2, 1, -1) & \delta(q_2, 2, \lambda) : (q_0, 2, 1)
\end{array}
$$

Dostáváme konfigurační graf



a označené cesty

$$
\begin{array}{llll}
(q_0, .) & \xrightarrow{0} & (\lambda, 2.), & M_\lambda(0) = \lambda \\
(q_0, .) & \xrightarrow{10} & (\lambda, 2.0), & M_\lambda(10) = 0 \\
(q_0, .) & \xrightarrow{110} & (\lambda, 2.1), & M_\lambda(110) = 1 \\
(q_0, .) & \xrightarrow{1110} & (\lambda, 2.00), & M_\lambda(1110) = 00 \\
(q_0, .) & \xrightarrow{1110} & (\lambda, 2.10), & M_\lambda(11110) = 10
\end{array}
$$

Teorii algoritmické složitosti lze rozvíjet abstraktně, v rámci teorie částečně rekursivních funkcí. Turingův automat nad abecedou A určuje pro každé $u \in A^*$ sebeomezující částečně rekursivní funkci $M_u : \mathcal{D}(M_u) \to A^*$, kde $\mathcal{D}(M_u) \subseteq B^*$, pro kterou platí

$$p, q \in \mathcal{D}(M_u) \implies p \not\sqsubseteq_p q$$

O abecedách A, B předpokládáme, že jsou lineárně uspořádáné a tato uspořádání rozšiřujeme na lexikografické uspořádání na celém A^*, při kterém kratší slova předchází delší slova.

Definice 29 *Nechť M je Turingův automat nad A. Kanonický program slova $u \in A^*$ a jeho složitost je*

$$P_M(u) = \min\{p \in B^* : \ M_\lambda(p) = u\}, \ \ K_M(u) = |P_M(u)|.$$

Pro $n \in \mathbb{N}$ je $P_M(n)$ a $K_M(n)$ kanonický program a složitost binárního zápisu $b(n)$ čísla n.

Kanonický program je tedy první program který u generuje a složitost u je délka tohoto kanonického programu. Pokud takový program neexistuje, není $P_M(u)$ definováno a $K_M(u) = \infty$. Přechodovou funkci δ Turingova automatu lze kódovat binárním slovem $D(M)$ tak, že množina všech $D(M)$ tvoří prefixový kód. Popis automatu M s přechodovou funkcí δ můžeme definovat jako seznam

$$D(M) = q_1 a_1 b_1 q_1' a_1' z_1, \ldots, q_m a_m b_m q_m' a_m' z_m.$$

kde $\delta(q_i, a_i, b_i) = (q_i', a_i', z_i)$. Prvky abeced $\overline{A}$, $\overline{B}$, $\overline{Q}$ a $\{-1, 0, 1\}$ jsou kódovány binárně. Tyto binární kódy jsou odděleny čárkami a celý seznam je ukončen tečkou. Takový seznam je tedy slovo čtyřprvkové abecedy $\{0, 1, ,, .\}$. Kódujeme-li písmena této abecedy binárními slovy délky 2, dostáváme popis $D(M)$ jako slovo binární abecedy.

Definice 30 *Říkáme, že automat U je univerzální nad abecedou A, pokud existuje funkce popisu D, která každému automatu M nad A přiřazuje slovo $D(M) \in B^*$ a platí*

(1) Pokud $M \neq M'$, pak $D(M) \not\sqsubseteq_p D(M')$

(2) Pro každé $p \in B^*$, $u \in A^*$ a každý automat M platí $M_u(p) = U_u(D(M)p)$, kde levá strana je definována právě když je definována pravá strana.

V teorii automatů se dokazuje, že pro každou abecedu A existuje funkce popisu a univerzální automat nad A.

Tvrzení 83 *Je-li U univerzální automat, pak pro každý automat M existuje konstanta C_M taková, že pro každé $x \in A^*$ platí $K_U(x) \leq K_M(x) + C_M$. Je-li V také univerzální automat, pak existuje konstanta C_{UV} taková, že*

$$\forall x \in A^*, |K_U(x) - K_V(x)| < C_{UV}.$$

Důkaz: Zřejmě $C_M = |D(M)|$. □

Je-li U univerzální, je $P_U : A^* \to B^*$ definováno pro každé $x \in A^*$ a je to prefixový kód. Pokud $P_U(x) \sqsubseteq_p P_U(y)$, pak $x = y$. Funkce P_U ovšem není rekursivní. Ukážeme nyní že existují algoritmicky složitá slova, která jsou sama svým nejkratším popisem. Algoritmicky jednoduchých slov totiž nemůže být příliš mnoho, protože je málo krátkých slov.

Tvrzení 84 *Nechť U je univerzální automat nad abecedou A.*

(1) $|\{x \in A^* : K_U(x) \leq n\}| \leq 2^n$ pro každé $n \in \mathbb{N}$.

(2) $|\{x \in A^* : K_U(x) < a\}| < 2^a$ pro každé $a \geq 0$ reálné.

(3) Pro každé $n \geq 0$ existuje $x \in A^n$, pro které $K_U(x) \geq n \cdot \log |A|$.

Důkaz: . (1) Zobrazení $x \mapsto P_U(x)0^{n-|P_U(x)|}$ z množiny $\{x \in A^* : K_U(x) \leq n\}$ do B^n je prosté, protože žádný prefix programu není program.
(2) Pro $a = 0$ je $0 = |\{x \in A^* : K_U(x) < 0\}| < 1 = 2^a$. Pro $a > 0$ existuje jediné $n \in \mathbb{N}$ pro které $a - 1 \leq n < a$ a platí

$$|\{x \in A^* : K_U(x) < a\}| = |\{x \in A^* : K_U(x) \leq n\}| \leq 2^n < 2^a$$

(3) Podle (1) platí $|\{x \in A^* : K_U(x) < n \cdot \log |A|\}| < 2^{n \log |A|} = |A^n|$, takže množina $A^n \setminus \{x \in A^* : K_U(x) < n \cdot |A|\}$ je neprázdná. □

Tvrzení 85 *Nechť U je univerzální automat nad abecedou A. Pak existuje C takové že*

(1) $\forall n \in \mathbb{N}, K_U(n) \leq \log(n+1) + 2 \log \log(n+2) + C$.

(2) $\forall u \in A^*, K_U(u) \leq |u| \log |A| + 4 \log(|u| + 1) + C$.

(3) $\forall u \in A^*, K_U(u) \leq |u| \cdot \mathcal{H}(\mathfrak{P}_u) + 2(|A| + 1)(\log(|u| + 1) + C$.

Důkaz plyne z Tvrzení 67, 68 a Věty 77, neboť příslušné kódy zde sestrojené jsou zřejmě rekursivní. Z Věty 73 pak bezprostředně plyne

Věta 86 *Nechť U je univerzální automat nad A a $X = (X_i)_{i \geq 0}$ bernoulliovský proces nad A s rozdělením $P \in \Delta(A)$. Pak*

$$\lim_{n \to \infty} \frac{K_U(X_{[0,n)})}{n} = \mathcal{H}(P) \textsf{ s.j.}$$

4.5 Pravděpodobnost zastavení

Představme si, že na vstupní pásce univerzálního Turingova automatu je náhodný program. S jakou pravděpodobností takový program vypočítá dané slovo výstupní abecedy a s jakou pravděpodobostí vůbec něco vypočítá?

Definice 31 *Nechť U je univerzální Turingův automat nad A. Absolutní pravděpodobnost $\omega(u)$ slova $u \in A^*$ a pravděpodobnost zastavení Ω (Chaitinovo číslo) je*

$$\omega_U(u) = \sum_{p \in U_\lambda^{-1}(u)} 2^{-|p|}, \quad \Omega_U = \sum_{u \in A^*} \omega_U(u) = \sum_{p \in \mathcal{D}(U_\lambda)} 2^{-|p|}$$

Každá konečná množina $M \subseteq \mathcal{D}(U)$ je prefixový kód, takže podle Kraftovy nerovnosti platí $\sum_{p \in M} 2^{-|p|} \leq 1$. Je tedy také $0 < \omega_U(x) < \Omega_U \leq 1$.

Věta 87 *Existuje C takové že pro všechna $u \in A^*$*

$$K_U(u) - C \leq -\log \omega_U(u) \leq K_U(u)$$

Důkaz: Z definice $\omega_U(u)$ bezprostředně plyne $2^{-K_U(u)} \leq \omega_U(u)$, tedy $-\log \omega_U(u) \leq K_U(u)$. Pro důkaz opačné nerovnosti sestrojíme Turingův automat M pro který platí $\omega_U(u) \leq 2^{-K_M(u)+2}$. Existuje prostá rekursivní posloupnost dvojic $(u_i, n_i)_{i \in \mathbb{N}}$ která obsahuje všechny dvojice $(u, n) \in B^* \times \mathbb{N}$ takové, že $\omega_U(u) > 2^{-n+1}$. Tuto posloupnost lze získat tak že postupně (a současně) počítáme všechny hodnoty $\omega_U(u)$: Simulujeme kroky všech programů $p \in B^+$ a pokud se p zastaví s výsledkem $U_\lambda(p) = u$, připočítáme k $\omega_U(u)$ číslo $2^{-|p|}$. Pro pevné $u \in B^*$ položme $N_u = \min\{n_i : u_i = u\}$, $I_u = \{i \in \mathbb{N} : u_i = u\}$. Pak

$$\sum_{i \in I_u} 2^{-n_i} \leq 2^{-N_u} + 2^{-N_u - 1} + \cdots \leq 2^{-N_u+1} < \omega_U(u) < 1$$

takže existuje prefixový kód $\{p_i \in B^+ : i \in I_u\}$, pro který $|p_i| = n_i$. Existuje tedy rekursivní posloupnost dvojic (u_i, p_i), kde $p_i \in B^+$, $|p_i| = n_i$ a pro každé u je $\{p_i : i \in I_u\}$ prefixový kód. Automat M při vstupu p prochází toto posloupnost dokud nenajde první index i pro který platí $p_i = p$. V tomto případě vypíše u_i a zastaví se. Je-li tedy $N_u = n_i$, je $M_\lambda(p_i) = u_i$ a $K_M(u) = |p_i| = N_u$, takže $\omega_U(u) \leq 2^{-N_u+2} = 2^{-K_M(u)+2}$. Odtud $-\log \omega_U(u) \geq K_M(u) - 2 \geq K_U(u) - 2 - C_M$. □

Pravděpodobnost zastavení má několik pozoruhodných vlastností. Vyjádříme-li ji v binárním tvaru, má každý její počáteční úsek délky n algoritmickou složitost nejméně $n - C$ a všechna binární slova se v ní vyskytují se stejnou pravděpodobností. Pišme Ω v binárním tvaru jako

$$\Omega = \sum_{n=0}^{\infty} \Omega_i \cdot 2^{-i-1}, \quad \forall i \in \mathbb{N}, \exists j \geq i, \Omega_j = 0$$

Tvrzení 88 *Existuje Turingův automat, který na základě vstupu $\mathfrak{b}_{\mathbb{N}}(n)\Omega_{[0,n)}$ vypíše seznam všech slov složitosti nejvýše n, tedy právě prvky množiny $\{x \in A^* : K_U(x) \leq n\}$.*

Důkaz: Uvažujme automat

Omega(x:real; n:integer);
    $y := 0$; $list := (q_0, \lambda, \lambda, \lambda)$;
    while $y \leq x$ do begin
        for all (q, v, u, p) in $list$ do begin
            for all $(q, v, u) \xrightarrow{b} (q', v', u')$ add (q', v', u', pb) to $list$;
            remove (q, v, u, p) from $list$;
            end;
        for all (q, v, u, p) in $list$ with $q = \lambda$ do begin
            $y := y + 2^{-|p|}$;
            if $|p| \leq n$ then write(u,',');
            remove (q, v, u, p) from $list$;
            end;
        end;
    write('.');
    end;

Automat současně počítá všechny programy a přitom sčítá pravděpodobnosti těch, které se zastaví. Současně na výstup zapíše slovo, který takový program vypočítal. Automat se zastaví pokud součet těchto pravděpodobností v proměnné y překročí danou mez x. Proměnná $list$ obsahuje seznam konfigurací všech počítaných programů. Pokud je na vstupu $x = 1$ a $n = \infty$, program se nikdy nezastaví a hodnota proměnné y přitom konverguje k Ω. Pokud je na vstupu $n < \infty$ a $x = \Omega_{[0,n)}$, program se po konečném počtu kroků zastaví a hodnota y pak splňuje $\Omega_{[0,n)} < y \leq \Omega$. Protože $\Omega - \Omega_{[0,n)} \leq 2^{-n}$, je také $\Omega - y < 2^{-n}$ a žádný program délky nejvýše n (kromě těch které byly nalezeny) se již nezastaví. $\square$

Věta 89 *Existuje $C > 0$ takové, že $K(\Omega_{[0,n)}) > n - C$ pro všechna $n \in \mathbb{N}$.*

Důkaz: Existuje Turingův automat M, který na základě kanonického programu $P_U(\Omega_{[0,n)})$ vypočítá první slovo $u \in A^*$ složitosti větší než n. Automat nejprve vypočítá n a $\Omega_{[0,n)}$ a spustí program z Tvrzení 88. Nakonec nalezne první slovo, které se nenalézá na pracovní pásce. To je první slovo složitosti větší než n. Je tedy

$$M(P_U(\Omega_{[0,n)})) = \min\{x \in A^* : \ K_U(x) > n\} = x^{(n)} \in A^*$$

$$K_U(\Omega_{[0,n)}) = |P_U(\Omega_{[0,n)})| \geq K_M(x^{(n)}) \geq K_U(x^{(n)}) - C_M > n - C_M \ \square$$

Věta 90 *Binární rozvoj Ω obsahuje všechna binární slova se stejnou frekvencí, tj.*

$$\forall u \in B^*, \lim_{n\to\infty} |\{i < n : \ \Omega_{[i,i+|u|)} = u\}|/n = 2^{-|u|}$$

Důkaz: Pro dané $k > 0$ označme $\Omega^{[k]} = (\Omega_{[ki,ki+k)})_{i\geq 0}$ nekonečné slovo v abecedě B^k a $\Omega^{[k]}_{[0,n)}$ jeho prefix délky n (který odpovídá prefixu Ω délky kn). Podle Věty 90 platí

$$kn - C \leq K(\Omega^{[k]}_{[0,n)}) \leq n \cdot \mathcal{H}(\mathfrak{P}_{\Omega^{[k]}_{[0,n)}}) + 2(|A^k| + 1)\log(n+1)$$

$$k \leq \liminf_{n\to\infty} \frac{1}{n}\mathcal{H}(\mathfrak{P}_{\Omega^{[k]}_{[0,n)}}) \leq \limsup_{n\to\infty} \frac{1}{n}\mathcal{H}(\mathfrak{P}_{\Omega^{[k]}_{[0,n)}}) \leq k$$

Je tedy $\lim_{n\to\infty} \mathcal{H}(\mathfrak{P}_{\Omega^{[k]}_{[0,n)}}) = k$, takže $\lim_{n\to\infty} \mathfrak{P}_{\Omega^{[k]}_{[0,n)}}(u) = 2^{-k}$. Stejný argument použijeme na posunutá nekonečná binární slova $\sigma^j(\Omega) = (\Omega_{i+j})_{i\geq 0}$ pro $0 \leq j < k$ a z kombinací těchto výsledků již plyne požadovaná rovnost. $\square$

Pokud se nějaký program $p \in B^*$ nezastaví, tj. pokud $p \notin \mathcal{D}(U)$, můžeme se pokusit dokázat tento fakt matematickými prostředky. Takové důkazy lze formalizovat v nějaké teorii prvního řádu, například v Peanově aritmetice, nebo v nějakém jejím rozšíření. Důkaz je posloupnost formulí, $\varphi_0\varphi_1\ldots\varphi_n$, kde každá formule je buď axiom, nebo vyplývá z předcházejících formulí podle dedukčních pravidel. Existuje algoritmus, který ověřuje zda daná posloupnost formulí je či není axiom. Uvažujme různé teorie T s různě silnými axiomatikami (ale se stejným jazykem). Předpokládáme, že tyto teorie jsou rekursivně axiomatizovatelné, tj. existuje algoritmus, který pro každou formuli rozhodne, zda je či není axiom. Pak můžeme sestrojit obecný algoritmus, který na základě popisu axiomatického systému $\alpha = D(T) \in B^*$ teorie T prochází v nekonečném cyklu postupně všechny konečné posloupnosti formulí, ověřuje zda to jsou důkazy a pokud ano, vypíše na výstup jejich poslední člen. Takový algoritmus tedy postupně vypisuje celou (rekursivně spočetnou) množinu dokazatelných formulí. Při studiu problému zastavení se omezíme na formule tvaru $\Omega_i = 0$ a $\Omega_i = 1$. Zajímá-li nás počáteční úsek $\Omega_{[0,n)}$, necháme algoritmus vypisovat jen dokazatelné formule $\Omega_i = c_i$ pro $i < n$.

Tvrzení 91 *Existuje automat M, který na základě vstupu $n \in \mathbb{N}$ a popisu axiomatického systému $\alpha = D(T)$ vypisuje postupně tvrzení tvaru $\Omega_i = c_i$, která jsou dokazatelná v teorii T. Automat se zastaví právě když pro každé $i < n$ najde jednu z vět $\Omega_i = 0$ nebo $\Omega_i = 1$.*

Tato konstrukce vede na alternativní důkaz Gödelovy věty o neúplnosti o existenci nerozhodnutelných vět:

Věta 92 *V každém bezesporném rekursivně axiomatizovatelném rozšíření Peanovy aritmetiky existuje tvrzení, které v ní není ani dokazatelné ani vyvratitelné.*

Důkaz: Pokud se program M z Tvrzení 91 zastaví při vstupu $\alpha = D(M)$ a $n \in \mathbb{N}$, pak platí

$$n - C_1 \leq K_U(\Omega_{[0,n)}) \leq |D(M)| + |\alpha| + 2\log n + C_2$$

Při pevném α je tato nerovnost splněna jen pro konečný počet čísel, takže existuje n, pro které se program M nezastaví. To znamená že existuje $i < n$, pro které tvrzení $\Omega_i = 0$ není rozhodnutelné v teorii T. $\square$

5 Ergodické procesy

Na základě zákonu velkých čísel jsme odvodili limitní věty teorie informace pro bernoulliovské procesy. Tyto věty lze zobecnit na širší třídu **ergodických** procesů, neplatí však obecně pro všechny stacionární procesy. To si ukážeme na příkladu konvexní kombinace bernoulliovských procesů.

Příklad 13 *Nechť* $(X_i : \Omega \to B)_{i\geq 0}$, $(Y_i : \Omega \to B)_{i\geq 0}$ *jsou bernoulliovské procesy v binární abecedě s rozděleními* $P = (p, 1-p)$, $Q = (q, 1-q)$, *a nechť* $V : \Omega \to B$ *je náhodná veličina s rozdělením* $P_V = (a, 1-a)$. *Předpokládejme dále, že všechny náhodné veličiny* X_i, Y_j, V *jsou navzájem nezávislé a* $0 < p, q, a < 1$. *Definujme proces* $(Z_i : \Omega \to B)_{i\geq 0}$ *předpisem*

$$Z_i(\omega) = \begin{cases} X_i(\omega) & \text{pro} \quad V(\omega) = 0 \\ Y_i(\omega) & \text{pro} \quad V(\omega) = 1 \end{cases}$$

Pak rozdělení procesu Z je konvexní kombinace rozdělení procesu X a Y:

$$\mathcal{P}_Z = a \cdot \mathcal{P}_X + (1-a) \cdot \mathcal{P}_Y$$

Pro střední hodnotu výskytu písmene $0 \in B$ platí

$$\mathbb{E}(|Z_i|_0) = a \cdot \mathbb{E}(|X_i|_0) + (1-a) \cdot \mathbb{E}(|Y_i|_0) = ap + (1-a)q$$

Dále platí

$$V(\omega) = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} \sum_{i<n} |Z_i(\omega)|_0/n = p$$

$$V(\omega) = 1 \quad \Longrightarrow \quad \lim_{n\to\infty} \sum_{i<n} |Z_i(\omega)|_0/n = q$$

To znamená

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i<n} |Z_i|_0 = p(1-V) + qV \textsf{ s.j.}$$

Náhodná veličina $|Z_{[0,n)}|_0/n$ nekonverguje skoro jistě ke ke konstantě ale k náhodné veličině $p + (q-p)V$. Střední hodnota této náhodné veličiny je ovšem

$$\mathbb{E}(p + (q-p)V) = pa + q(1-a) = \mathbb{E}(|Z_i|_0)$$

Ergodická věta se zabývá konvergencí funkcí stacionárních procesů. Tyto procesy je výhodné chápat jako dynamické systémy. Náhodný proces X můžeme chápat jako proces na pravděpodobnostním prostoru $(A^{\mathbb{N}}, \mathcal{B}, \mu)$, kde $\mathcal{B}$ je σ-algebra borelovských množin, a $\mu = \mathcal{P}_X$ je symbolická míra. Zobrazení posunu $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ zapomíná nultý člen sekvence x:

$$\sigma(x)_i = x_{i+1}$$

Tvrzení 93 *Symbolická míra je stacionární právě když je invariantní vůči* σ, *tj.*

$$\forall V \in \mathcal{B}, \mu(\sigma^{-1}(V)) = \mu(V)$$

Důkaz: Symbolická míra ν definovaná jako $\nu(V) = \mu(\sigma^{-1}(V))$ je rozložením posunutého procesu $(X_i)_{i\geq 1}$. $\square$

Zákon velkých čísel říká, že pro bernoulliovský proces X pro každé $u \in A^*$ platí

$$\lim_{n\to\infty} \frac{1}{n} |\{i < n : \ X_{[i,i+|u|)} = u\}| = P_X(u) \textsf{ s.j.}$$

Označíme-li $\chi_{[u]}$ charakteristickou funkci cylindru $[u]$, můžeme to psát ve tvaru

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} \chi_{[u]}(\sigma^i(x)) = P_X(u)$$

Ergodická věta se zabývá existencí limit tohoto tvaru. Přitom místo charakteristických funkcí uvažuje obecněji integrovatelné funkce.

5.1 Integrace

Uvažujme pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mu)$. Funkce $f : \Omega \to \mathbb{R}$ je měřitelná, jestliže $f^{-1}[a, b) \in \mathcal{A}$ pro každý interval $[a, b) \subseteq \mathbb{R}$. Měřitelná funkce je schodovitá, pokud její obor hodnot je konečný. V takovém případě je f lineární kombinace charakteristických funkcí: Existuje konečný soubor disjunktních měřitelných množin $(V_i)_{i\in I}$ a reálných čísel $(c_i)_{i\in I}$ takových že $f = \sum_{i\in I} c_i \cdot \chi_{V_i}$. Integrál schodovité funkce je

$$\int f\, d\mu = \sum_{i\in I} c_i \cdot \mu(V_i)$$

Integrál měřitelné nezáporné funkce $f : A^{\mathbb{N}} \to [0, \infty)$ definujeme jako supremum integrálů všech schodovitých funkcí menších než daná funkce:

$$\mathbb{E}(f) = \int f\, d\mu = \sup \left\{ \int s\, d\mu :\ s \text{ je schodovitá funkce a } s \le f \right\}$$

Integrál $\mathbb{E}(f) \in [0, \infty]$ může být nekonečný. Pro měřitelnou funkci $f : A^{\mathbb{N}} \to \mathbb{R}$ definujeme její kladnou a zápornou komponentu

$$f^+(x) = \max\{0, f(x)\}, \quad f^-(x) = \max\{0, -f(x)\}$$

takže f^+ a f^- jsou nezáporné měřitelné funkce a $f = f^+ - f^-$. Funkce f je integrovatelná, jestliže oba integrály $\int f^+ d\mu$ a $\int f^- d\mu$ jsou konečné. V tomto případě klademe

$$\mathbb{E}(f) = \int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

Množinu všech integrovatelných funkcí značíme $\mathcal{L}(\Omega, \mathcal{A}, \mu)$ nebo kratčeji $\mathcal{L}(\mu)$. Integrál funkce f přes měřitelnou množinu $V \in \mathcal{A}$ definujeme

$$\mathbb{E}_V(f) = \int_V f\, d\mu = \int f \cdot \chi_V\, d\mu$$

Příklad 14 *Nechť μ je bernoulliovská míra. Funkce $R(x) = \inf\{n > 0 :\ x_n = x_0\}$ je integrovatelná na $A^{\mathbb{N}}$, $\mathbb{E}_{[a]}(R) = 1$ a $\mathbb{E}(R) = |A|$.*

Důkaz: Funkce je konstantní na cylindrech $[u]$ takových že $u_0 = u_{|u|-1}$ a $u_i \ne u_0$ pro $0 < i < |u| - 1$. Pro $x \in [u]$ je $f(x) = |u| - 1$. Odtud

$$\begin{aligned}
\mathbb{E}_{[a]}(R) &= \sum_{n=0}^{\infty} \sum_{u\in(B\setminus\{a\})^n} (n+1)\cdot\mu(aua) = \sum_{n=0}^{\infty}(n+1)P(a)^2(1-P(a))^n \\
&= P(a)^2(\sum_{n\ge 0}(1-P(a))^n + \sum_{n\ge 1}(1-P(a))^n + \cdots) \\
&= P(a)(1 + (1-P(a) + (1-P(a))^2 + \cdots) = 1 \\
\mathbb{E}(R) &= \sum_{a\in A} \mathbb{E}_{[a]}(R) = |A|
\end{aligned}$$

5.2 Ergodická věta

Definice 32 *Dynamický systém je trojice $(\Omega, \mathcal{A}, T)$, kde $(\Omega, \mathcal{A})$ je měřitelný prostor a $T : \Omega \to \Omega$ je měřitelné zobrazení. Pravděpodobnostní míra $\mu : \mathcal{A} \to [0, 1]$ je T-invariantní, pokud $T\mu = \mu$, tj. $\mu(T^{-1}(V)) = \mu(V)$ pro každou $V \in \mathcal{A}$.*

Pro dynamický systém $(\Omega, \mathcal{A}, T)$, T-invariantní míru μ a $f \in \mathcal{L}(\mu)$ položme

$$\begin{aligned} s_n(f, x) &= \sum_{k=0}^{n-1} f(T^k(x)), \quad n \geq 1, x \in \Omega \\ V(f) &= \{x \in \Omega : \sup_{n>0} s_n(f, x) > 0\} \end{aligned}$$

Lemma 94 *Nechť μ je T-invariantní míra a $f : \Omega \to \mathbb{R}$ μ-integrovatelná funkce. Pak*

$$\int_{V(f)} f(x)\, d\mu \geq 0.$$

Důkaz: : Položme

$$\varphi_n(x) = \max\{0, s_1(f, x), \dots, s_n(f, x)\},\ V_n = \{x \in \Omega\ \ \varphi_n(x) > 0\}$$

Protože $f(x) + s_n(f, T(x)) = s_{n+1}(f, x)$,

$$\begin{aligned} f(x) + \varphi_n(T(x)) &= \max\{s_1(f, x), s_2(f, x), \dots, s_n(f, x), s_{n+1}(f, x)\} \\ &\geq \max\{s_1(f, x), s_2(f, x), \dots, s_n(f, x)\} \end{aligned}$$

Pro $x \in V_n$ tedy platí $f(x) + \varphi_n(T(x)) \geq \varphi_n(x)$. Protože $\varphi_n(x) \geq 0$, a $\varphi_n(x) = 0$ pro $x \notin V_n$, dostáváme

$$\int_{V_n} \varphi_n(T(x))\, d\mu \leq \int_\Omega \varphi_n(T(x))\, d\mu, \quad \int_{V_n} \varphi_n(x)\, d\mu = \int_\Omega \varphi_n(x)\, d\mu$$

Odtud integrací nerovnosti $f(x) \geq \varphi_n(x) - \varphi_n(Tx)$

$$\int_{V_n} f(x)\, d\mu \geq \int_{V_n} \varphi_n(x)\, d\mu - \int_{V_n} \varphi_n(T(x))\, d\mu \geq \int_\Omega \varphi_n(x)\, d\mu - \int_\Omega \varphi_n(T(x))\, d\mu = 0$$

Protože $V_n \subseteq V_{n+1}$ a $V(f) = \bigcup_{n>0} V_n$, platí $\int_{V(f)} f(x)\, d\mu \geq 0$. □

Věta 95 (Birkhoffova ergodická věta) *Nechť $(\Omega, \mathcal{A}, T)$ je dynamický systém, μ je T-invariantní míra a $f : \Omega \to \mathbb{R}$ μ-integrovatelná funkce. Pak existuje μ-integrovatelná funkce $f^* : \Omega \to \mathbb{R}$ taková že $f^*(T(x)) = f^*(x)$* **s.j.** *a platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = f^*(x) \text{ s.j. }, \quad \int f^*(x)\, d\mu = \int f(x)\, d\mu.$$

Důkaz: : Pro reálná $a < b$ označme

$$E = \left\{ x \in \Omega : \liminf_{n\to\infty} \frac{s_n(f, x)}{n} < a < b < \limsup_{n\to\infty} \frac{s_n(f, x)}{n} \right\}$$

Protože $s_n(f, T(x)) - s_n(f, x) = f(T^n(x)) - f(x)$, je množina E T-invariantní, tj. $T^{-1}(E) = E$. Definujme integrovatelné funkci $g, h : X \to \mathbb{R}$ předpisem

$$g(x) = \begin{cases} f(x) - b & \text{pro } x \in E \\ 0 & \text{pro } x \notin E \end{cases}, \quad h(x) = \begin{cases} a - f(x) & \text{pro } x \in E \\ 0 & \text{pro } x \notin E \end{cases}$$

Ukážeme že $V(g) = E = V(h)$. Pokud $x \in E$, pak $T^n(x) \in E$ pro všechna n, takže

$$\begin{aligned}
\limsup_{n\to\infty} \frac{s_n(g,x)}{n} &= \limsup_{n\to\infty} \frac{s_n(f,x)}{n} - b > 0 \\
\limsup_{n\to\infty} \frac{s_n(h,x)}{n} &= a - \liminf_{n\to\infty} \frac{s_n(f,x)}{n} > 0
\end{aligned}$$

Odtud $\sup_{n>0} s_n(g,x) > 0$ takže $x \in V(g)$, a podobně $\inf_{n>0} s_n(h,x) > 0$, takže $x \in V(h)$. Pokud $x \notin E$, pak pro každé n, $T^n(x) \notin E$, takže $g(T^n x) = h(T^n x) = 0$. Odtud $s_n(g,x) = s_n(h,x) = 0$, a tedy $x \notin V(g)$ a $x \notin V(h)$. Je tedy $V(g) = E = V(h)$. Podle Lemma 94

$$\begin{aligned}
b\mu(E) &\le \int_{V(g)} g\, d\mu + b\mu(E) = \int_E g\, d\mu + \int_E b\, d\mu = \int_E f\, d\mu \\
&= \int_E a\, d\mu - \int_E h\, d\mu = a\mu(E) - \int_{V(h)} h\, d\mu \le a\mu(E)
\end{aligned}$$

takže $\mu(E) = 0$. Protože to platí pro všechna $a < b$ existuje $f^*(x) = \lim_{n\to\infty} s_n(f,x)/n$ skoro všude. Pokud $f^*(x)$ existuje, existuje i $f^*(T(x)) = f^*(x)$, takže f^* je skoro všude invariantní. Pro každé $n > 0$ je

$$\int \left| \frac{s_n(f,x)}{n} \right| d\mu \le \frac{1}{n} \int \sum_{k<n} |f(T^k x)|\, d\mu = \frac{1}{n} \sum_{k<n} \int |f(T^k(x))|\, d\mu = \int |f(x)|\, d\mu < \infty$$

takže f^* je integrovatelná. Pro pevné $q > 0$ a $p \in \mathbb{Z}$ označme

$$C_p = \left\{ x \in \Omega;\ \frac{p}{q} < \liminf_{n\to\infty} \frac{s_n(f,x)}{n} \ \&\ \limsup_{n\to\infty} \frac{s_n(f,x)}{n} < \frac{p+1}{q} \right\}$$

$$g_p(x) = \begin{cases} f(x) - \frac{p}{q} & \text{pro } x \in C_p \\ 0 & \text{pro } x \notin C_p \end{cases}, \quad h_p(x) = \begin{cases} \frac{p+1}{q} - f(x) & \text{pro } x \in C_p \\ 0 & \text{pro } x \notin C_p \end{cases}$$

takže $V(g_p) = C_p = V(h_p)$ a podle Lemma 94 platí

$$\frac{p}{q} \cdot \mu(C_p) \le \int_{C_p} f\, d\mu \le \frac{p+1}{q} \cdot \mu(C_p)$$

Zřejmě také

$$\frac{p}{q} \cdot \mu(C_p) \le \int_{C_p} f^*\, d\mu \le \frac{p+1}{q} \cdot \mu(C_p)$$

takže

$$\begin{aligned}
\left| \int f^*(x)\, d\mu - \int f(x)\, d\mu \right| &= \left| \sum_{p=-\infty}^{\infty} \int_{C_p} f^*\, d\mu - \int_{C_p} f\, d\mu \right| \\
&\le \sum_{p=-\infty}^{+\infty} \left| \int_{C_p} (f^*(x) - f(x))\, d\mu \right| \le \sum_{p=-\infty}^{+\infty} \frac{\mu(C_p)}{q} \le \frac{1}{q}
\end{aligned}$$

Protože to platí pro každé q, tyto dva integrály se rovnají. □

Ergodická věta říká že funkce f^* je invariantní, tj. $f = f \circ T$ **s.j.** . V případě že f^* je skoro jistě konstantní, říkáme že dynamický systém je ergodický.

Definice 33 *Nechť* $(\Omega, \mathcal{A}, T)$ *je dynamický systém.* T*-invariantní míra* μ *je* **ergodická**, *pokud každá skoro všude invariantní funkce je skoro všude konstantní, tj.*

$$f \in \mathcal{L}(\mu), f \circ T = f \ \textbf{s.j.} \implies \exists c \in \mathbb{R}, f = c \ \textbf{s.j.}$$

Tvrzení 96 *Nechť* $(\Omega, \mathcal{A}, T)$ *je dynamický systém a* μ *je* T*-invariantní míra. Následující podmínky jsou ekvivalentní:*

(1) μ *je ergodická.*

(2) Každá invariantní množina má míru nula nebo jedna, tj.

$$U \in \mathcal{A} \ \& \ T^{-1}(U) = U \implies \mu(U) \in \{0, 1\}$$

(3) Každá subinvariantní množina má míru nula nebo jedna, tj.

$$U \in \mathcal{A} \ \& \ T^{-1}(U) \subseteq U \implies \mu(U) \in \{0, 1\}$$

(4) Každá superinvariantní množina má míru nula nebo jedna, tj.

$$U \in \mathcal{A} \ \& \ U \subseteq T^{-1}(U) \implies \mu(U) \in \{0, 1\}$$

(5) Každá skoro invariantní množina má míru nula nebo jedna, tj.

$$U \in \mathcal{A} \ \& \ \mu(T^{-1}(U) \triangle U) = 0 \implies \mu(U) \in \{0, 1\}]$$

(6) Každá subinvariantní integrovatelná funkce je skoro jistě konstantní

$$f \in \mathcal{L}(\mu) \ \& \ f \circ T \leq f \ \textsf{s.j.} \implies \exists c \in \mathbb{R}, f = c \ \textsf{s.j.}$$

Důkaz: (1) $\Rightarrow$ (2) : Je-li $T^{-1}(U) = U$, je χ_U invariantní a tedy skoro jistě konstantní. To znamená že buď $\chi_U = 0$ **s.j.** nebo $\chi_U = 1$ **s.j.** a tedy $\mu(U) \in \{0, 1\}$.
(2) $\Rightarrow$ (3). Položme $V = \bigcap_{n \geq 0} T^{-n}(U)$. Pak $V \in \mathcal{A}$ a $T^{-1}(V) = V$, takže $\mu(V) \in \{0, 1\}$. Protože $\mu(V) = \lim_{n \to \infty} \mu(T^{-n}(U)) = \mu(U)$, je $\mu(U) \in \{0, 1\}$.
(3) $\Leftrightarrow$ (4) : Platí $U \subseteq T^{-1}(U)$ právě když $T^{-1}(\Omega \setminus U) \subseteq \Omega \setminus U$ a $\mu(\Omega \setminus U) \in \{0, 1\}$ právě když $\mu(U) \in \{0, 1\}$.
(3) & (4) $\Rightarrow$ (5) : Položme $U_0 = \bigcap_{n \geq 0} T^{-n}(U)$, $U_1 = \bigcup_{n \geq 0} T^{-n}(U)$. Pak $U_0 \subseteq T^{-1}(U_0)$ a $T^{-1}(U_1) \subseteq U_1$, takže obě tyto množiny mají míru nula nebo jedna. Dále platí

$$\mu(U \triangle T^{-2}(U)) \leq \mu(U \triangle T^{-1}(U)) + \mu(T^{-1}(U \triangle T^{-1}(U)) = 0,$$

a podobně odvodíme $\mu(T^{-n}(U) \triangle T^{-k}(U)) = 0$ pro každé n, k. Odtud

$$\mu(U_1 \setminus U_0) = \mu\left(\bigcup_{n \geq 0} \bigcup_{k \geq 0} T^{-n}(U) \setminus T^{-k}(U)\right) \leq \sum_{n \geq 0} \sum_{k \geq 0} \mu(T^{-n}(U) \setminus T^{-k}(U)) = 0$$

takže $\mu(U_0) = \mu(U) = \mu(U_1) \in \{0, 1\}$.
(5) $\Rightarrow$ (1) : Pro každé $a < b$ je množina $V_{a,b} = f^{-1}(a, b)$ skoro invariantní, takže má míru nula nebo jedna. Existuje $n \in \mathbb{Z}$ takové že $\mu(V_{n,n+1}) = 1$ a dělením intervalu na intervaly poloviční délky dospějeme ke konstantě $c \in [n, n+1]$ takové že $f = c$ **s.j.** .
(3) $\Rightarrow$ (6) : Pro $d \in \mathbb{R}$ položme $V_d = \{x \in \Omega : \ f(x) \geq d\}$. Pak $T^{-1}(V_d) \leq V_d$, takže $\mu(V_d) \in \{0, 1\}$. Položme $c = \inf\{d \in \mathbb{R} : \ \mu(V_d) = 1\}$. Pak $f = c$ **s.j.**
(6) $\Rightarrow$ (1) plyne z definice. □

Shrnutí 97 *Je-li* $(\Omega, \mathcal{A}, T)$ *dynamický systém a* μ T*-ergodická míra, pak pro každou* $f \in \mathcal{L}(\mu)$ *platí*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i < n} f(T^n(\omega)) = \mathbb{E}(f) \ \textsf{s.j.}$$

Věta 98 *Nechť* $(\Omega, \mathcal{A}, T)$ *je dynamický systém a* μ *je* T*-invariantní míra. Následující podmínky jsou ekvivalentní:*

(1) μ je ergodická.

(2) Množinu kladné míry navštíví skoro každý bod, tj.

$$U \in \mathcal{A} \ \& \ \mu(U) > 0 \implies \mu\left(\bigcup_{n\geq 0} T^{-n}(U)\right) = 1$$

(3) Pro každé $U, V \in \mathcal{A}$ platí

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \mu(T^{-k}(U) \cap V) = \mu(U) \cdot \mu(V)$$

(4) μ je extremální, tj. není vlastní konvexní kombinací žádných dvou jiných měr.

Důkaz: (1) $\Leftrightarrow$ (2) plyne bezprostředně z Tvrzení 96
(1) $\Rightarrow$ (3) :

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \mu(T^{-k}(U) \cap V) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \int \chi_{T^{-k}(U)}(x) \cdot \chi_V(x)\, d\mu \\
&= \int \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \chi_U(T^k(x))\right) \cdot \chi_V(x)\, d\mu \\
&= \int \mu(U) \cdot \chi_V(x)\, d\mu = \mu(U) \cdot \mu(V)
\end{aligned}$$

(3) $\Rightarrow$ (1) : Nechť U je invariantní množina tj. $T^{-1}(U) = U$. Pak

$$\mu(U)^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \mu(T^{-k}(U) \cap U) = \mu(U)$$

takže $\mu(U) \in \{0, 1\}$.
(1) $\Rightarrow$ (4) : Předpokládejme sporem, že T-ergodická míra μ je vlastní lineární kombinací $\mu = a_0\nu_0 + a_1\nu_1$, kde $0 < a_0, a_1 < 1$ a $a_0 + a_1 = 1$. Protože $\nu_0 \neq \nu_1$, existuje $U \in \mathcal{A}$ taková že $\nu_0(U) \neq \nu_1(U)$. Předpokládejme bez újmy na obecnosti $\nu_0(U) < \nu_1(U)$. Označme V množinu těch $x \in \Omega$, pro které existuje limita

$$f(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \chi_U(T^k(x))$$

Podle Ergodické věty je $\nu_0(V) = \nu_1(V) = 1$ a platí

$$\int f(x)\, d\nu_0 = \nu_0(U), \quad \int f(x)\, d\nu_1 = \nu_1(U)$$

Zvolme b takové že $\nu_0(U) < b < \nu_1(U)$. Množina $W = \{x \in V : \ f(x) > b\}$ je μ-skoro invariantní. Předpokládejme že $\mu(W) \in \{0, 1\}$:

$$\mu(W) = 0 \quad \Rightarrow \quad \nu_1(W) = 0 \ \Rightarrow \ \nu_1(U) = \int f\, d\nu_1 \leq b$$

$$\mu(W) = 1 \quad \Rightarrow \quad \nu_0(W) = 1 \ \Rightarrow \ \nu_0(U) = \int f\, d\nu_0 \geq b$$

a v obou případech dostáváme spor. Míra μ tedy není ergodická.
$\neg(1) \Rightarrow \neg(4)$: Nechť U je T-invariantní množina (tj. $T^{-1}(U) = U$), pro kterou $0 < \mu(U) < 1$. Definujme míry ν_0, ν_1 předpisem

$$\nu_0(V) = \mu(V \cap U)/\mu(U), \quad \nu_1(V) = \mu(V \setminus U)/\mu(\Omega \setminus U)$$

Pak $\mu = \mu(U) \cdot \nu_0 + (1 - \mu(U)) \cdot \nu_1$. $\square$

Tvrzení 99 *Symbolická stacionární míra $\mu : A^* \to [0,1]$ je ergodická právě když pro každá slova $u, v \in A^*$ platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k<n} \sum_{w\in A^k} \mu(uwv) = \mu(u)\cdot\mu(v)$$

Každá bernoulliovská míra je ergodická. Markovská míra stochastické matice R je ergodická právě když R je ireducibilní. Ergodická věta umožňuje zobecnit některé limitní věty teorie informace.

5.3 Entropická věta

Horní a dolní limita posloupnosti množin $(V_n \subseteq \Omega)_{n\geq 0}$ je

$$\begin{aligned}
\liminf_{n\to\infty} V_n &= \bigcup_{k\geq 0}\bigcap_{n\geq k} V_n = \{x\in\Omega:\ \exists k, \forall n\geq k, x\in V_n\}\\
\limsup_{n\to\infty} V_n &= \bigcap_{k\geq 0}\bigcup_{n\geq k} V_n = \{x\in\Omega:\ \forall k, \exists n\geq k, x\in V_n\}
\end{aligned}$$

Zřejmě platí $\liminf_{n\to\infty} V_n \subseteq \limsup_{n\to\infty} V_n$ a $\Omega\setminus\liminf_{n\to\infty} V_n = \limsup_{n\to\infty}(\Omega\setminus V_n)$. Pro monotónní posloupnosti množin platí

$$\begin{aligned}
V_n\subseteq V_{n+1} &\implies \liminf_{n\to\infty} V_n = \limsup_{n\to\infty} V_n = \bigcup_{n\geq 0} V_n\\
V_{n+1}\subseteq V_n &\implies \liminf_{n\to\infty} V_n = \limsup_{n\to\infty} V_n = \bigcap_{n\geq 0} V_n
\end{aligned}$$

Tvrzení 100 (Borel-Cantelliho lemma) *Nechť U_n, V_n jsou posloupnosti měřitelných množin. Pak*

$$\begin{aligned}
\sum_{n\geq 0}\mu(V_n) < \infty &\implies \mu(\limsup_{n\to\infty} V_n) = 0\\
\mu(\liminf_{n\to\infty} U_n) = 1\ \&\ \mu(\limsup_{n\to\infty}(U_n\cap V_n)) = 0 &\implies \mu(\limsup_{n\to\infty} V_n) = 0
\end{aligned}$$

Důkaz: (1) Pro každé $\varepsilon > 0$ existuje k takové že $\sum_{n\geq k}\mu(V_n) < \varepsilon$, takže

$$\mu(\limsup_{n\to\infty} V_k) \leq \mu\left(\bigcup_{n\geq k} V_n\right) \leq \sum_{n\geq k}\mu(V_n) < \varepsilon$$

(2) Platí $(\liminf_{n\to\infty} U_n)\cap(\liminf_{n\to\infty}(\Omega\setminus(U_n\cap V_n))) \subseteq \liminf_{n\to\infty}(\Omega\setminus V_n)$. Obě množiny na levé straně mají míru 1, takže také jejich průnik má míru 1 a množina na pravé straně má také míru 1. □

Shrnutí 101 *Pro ergodickou míru μ a měřitelnou funkci f položme*

$$V_n(\varepsilon) = \left\{x\in\Omega:\ \left|\frac{1}{n}\sum_{i<n} f(T^n(x)) - \mathbb{E}(f)\right| < \varepsilon\right\}$$

Pak $\mu(\liminf_{n\to\infty} V_n(\varepsilon)) = 1$.

Definice 34 *Posloupnost intervalů $S = ([n_i, m_i))_{i<\ell(S)}$ je δ-rozklad intervalu $[0,k)$, pokud intervaly $[n_i, m_i)\subseteq[0,k)$ jsou navzájem disjunktní a $\sum_{i<\ell(S)} m_i - n_i \geq (1-\delta)k$.*
Slovo $u\in A^*$ je δ-pokryto množinou $\mathcal{T}\subseteq A^*$, pokud existuje δ-rozklad S intervalu $[0,|u|)$ takový že $u_{[n_i,m_i)}\in\mathcal{T}$ pro každé $i<\ell(S)$.

Lemma 102 *Nechť $h > 0$, nechť $M \subseteq A^*$ je množina slov taková že $|M \cap A^n| \leq 2^{nh}$ a nechť $D_n(\delta)$ je množina slov $u \in A^n$, která jsou δ-pokryta množinou M, tj.*

$$D_n(\delta) = \{u \in A^n : \exists([n_i, m_i))_{i<\ell}, m_i \leq n_{i+1}, u_{[n_i,m_i)} \in M, \sum_{i<\ell} m_i - n_i > (1-\delta)n\}$$

Pak $|D_n(\delta)| \leq 2^{n \cdot h} \cdot 2^{n \cdot \mathcal{H}(\delta,1-\delta)} \cdot |A|^{n\delta}$. Jestliže $n > 2L/\delta$, pak

$$\left\{u \in A^n : |\{i < n-L : \exists j < L, u_{[i,i+j)} \in M\}| > (1-\tfrac{\delta}{2})(n-L)\right\} \subseteq D_n(\delta)$$

Důkaz: δ-rozklad $S = ([n_i, m_i))_{i<\ell}$ je jednoznačně určen množinou $[0,n) \setminus \bigcup_{i<\ell}[n_i, m_i)$, která má nejvýše δn prvků. Počet takových množin je

$$\sum_{i<\delta \cdot n} \binom{n}{i} \leq 2^{n \cdot \mathcal{H}(\delta,1-\delta)}.$$

Pro pevný rozklad $S = ([n_i, m_i))_{i<\ell}$, počet všech slov u takových že $u_{[n_i,m_i)} \in M$ je nejvýše

$$|A|^{\delta \cdot n} \cdot \prod_{i<\ell} |M \cap A^{m_i - n_i}| \leq |A|^{\delta \cdot n} \cdot 2^{h \cdot n}$$

Předpokládejme že $|\{i < n-L : \exists j < L, u_{[i,i+j)} \in M\}| > (1-\frac{\delta}{2})(n-L)$ a sestrojme rozklad S. Položme $m_{-1} = 0$,

$$n_i = \min\{k \geq m_{i-1} : \exists j < L, u_{[k,k+j)} \in M\}, \quad m_i = \min\{j < L : u_{[n_i,j)} \in M\},$$

a $\ell(S)$ je první index i, pro který n_i není definováno. Pak

$$\begin{aligned} [0,n) \setminus \bigcup_{i<\ell}[n_i, m_i) &\subseteq \{i < n-L : \forall j < L, u_{[i,i+j)} \notin M\} \cup (n-L, L) \\ \left|[0,n) \setminus \bigcup_{i<\ell}[n_i, m_i)\right| &\leq \frac{\delta}{2}(n-L) + L \leq \delta \cdot n \ \square \end{aligned}$$

Věta 103 (Entropická Věta - Shannon,McMillan,Breiman) *Nechť $(X_i : \Omega \to A)_{i \geq 0}$ je ergodický proces. Pak*

$$\lim_{n\to\infty} \frac{\mathcal{I}_{X_{[0,n)}}(x)}{n} = \mathcal{H}(\mathcal{P}_X) \text{ s.j.}$$

Důkaz: Označme $\mu = \mathcal{P}_X$. Pro $x \in A^{\mathbb{N}}$ položme

$$h_n(x) = \frac{-\log \mu(x_{[0,n)})}{n}, \quad h(x) = \liminf_{n\to\infty} h_n(x).$$

Pak $h(\sigma(x)) \leq h(x)$ a podle Věty 98 existuje konstanta h, taková že $h(x) = h$ **s.j.** Ukážeme že $\limsup_{n\to\infty} h_n(x) = h$ **s.j.** Pro dané $\varepsilon > 0$ zvolme $\delta > 0$ takové že $\delta + \delta \cdot \log|A| + \mathcal{H}(\delta, 1-\delta) < \varepsilon$. Pro skoro všechna x platí $h_n(x) < h + \delta$ a tedy $\mu(x_{[0,n)}) \geq 2^{-n(h+\delta)}$ pro nekonečně mnoho n. Položme

$$M = \{u \in A^* : \mu(u) > 2^{-|u| \cdot (h+\delta)}\}$$

a nechť $D_n(\delta)$ je množina $u \in A^n$, která jsou δ-pokryta M. Pak $|M \cap A^n| < 2^{n(h+\delta)}$ a

$$|D_n(\delta)| \leq 2^{n \cdot (h+\delta)} \cdot 2^{n \cdot \mathcal{H}(\delta,1-\delta)} \cdot |A|^{n\delta} \leq 2^{n(h+\varepsilon)}$$

Protože $\mu\{x \in A^{\mathbb{N}} : \exists i, x_{[0,i)} \in M\} = 1$, existuje $L > 0$ takové že

$$\mu\{x \in A^{\mathbb{N}} : \exists i < L, x_{[0,i)} \in M\} > 1 - \frac{\delta}{4}$$

Položme

$$W_n = \{x \in A^{\mathbb{N}} : |\{i < n - L : \exists j < L, x_{[i,i+j)} \in M\}| > (1 - \tfrac{\delta}{2})(n - L)\}$$

Podle Shrnutí 101 je $\mu(\liminf_{n\to\infty} W_n)) = 1$ a podle Lemma 102 je $W_n \subseteq [D_n(\delta)]$. Položme

$$C_n = \{u \in A^n : \mu(u) < 2^{-n(h+2\varepsilon)}\}$$

Pak $\mu(C_n \cap D_n(\delta)) \leq |D_n(\delta)| \cdot 2^{-n(h+2\varepsilon)} \leq 2^{-n\varepsilon}$. Podle Borel-Cantelliho lemma je

$$\mu(\limsup_{n\to\infty}[C_n \cap D_n(\delta)]) = 0, \text{ takže } \mu(\limsup_{n\to\infty}[C_n]) = 0.$$

Pro $x \in A^{\mathbb{N}} \setminus \limsup_{n\to\infty}[C_n]$ platí $\limsup_{n\to\infty} h_n(x) \leq h+2\varepsilon$. Dokázali jsme tedy $\lim_{n\to\infty} h_n(x) = h$ **s.j.** Protože $\lim_{n\to\infty} \mathbb{E}(h_n) = \mathcal{H}(\mu)$, je $h = \mathcal{H}(\mu)$. □

5.4 Kódování ergodických procesů

Pro $n > k$, $m = \lceil n/k \rceil$, $u \in A^n$, $v \in A^k$ položme

$$|u|_v = |\{i < m : (uu)_{[ik,ik+k)} = v\}|, \quad \mathfrak{P}^k_u(v) = |u|_v / m$$

Pak $\mathfrak{P}^k_u \in \Delta(A^k)$.

Tvrzení 104 *Pro každý ergodický proces platí*

$$\lim_{n\to\infty} \mathcal{H}(\mathfrak{P}^k_{X_{[0,n}}) = \mathcal{H}(X_{[0,k)}) \textsf{ s.j.}$$

Důkaz: Proces $Y_i = X_{[ki,ki+k)}$ je ergodický a $\mathfrak{P}^k_{X_{[0,nk)}} = \mathfrak{P}_{Y_{[0,n)}}$. □

Věta 105 *Je-li X ergodický proces s rozdělením $\mathcal{P}_X = \mu$, pak existuje prefixový kód $\mathfrak{b}_\mu : A^* \to B^*$ takový že*

$$\lim_{n\to\infty} \frac{|\mathfrak{b}_\mu(X_{[0,n)})|}{n} = \mathcal{H}(X) \textsf{ s.j.}$$

Důkaz: Kód $\mathfrak{b}_\mu$ byl sestrojen v důkazu Věty 75 jako $\mathfrak{b}_\mu(u) = \mathfrak{b}_0(|u|) f_{|u|}(u)$, kde $f_n : A^n \to B^*$ je prefixový kód splňující $-\log \mu(u) \leq |f_n(u)| \leq -\log \mu(u) + 1$. Pro dané $\varepsilon > 0$ položme

$$V_n(\varepsilon) = \{x \in A^{\mathbb{N}} : n(\mathcal{H}(\mu) - \varepsilon) \leq -\log \mu(x_{[0,n)}) \leq n(\mathcal{H}(\mu) + \varepsilon)\}, \quad V(\varepsilon) = \liminf_{n\to\infty} V_n(\varepsilon)$$

Podle Entropické věty platí $\mu(V(\varepsilon)) = 1$. Pro $x \in V_n(\varepsilon)$ platí

$$n(\mathcal{H}(\mu) - \varepsilon) \leq |\mathfrak{b}_\mu(x_{[0,n)})| \leq n(\mathcal{H}(\mu) + \varepsilon) + 2\log(n+1) + 2$$

To znamená, že pro $x \in V(\varepsilon)$ platí

$$\mathcal{H}(\mu) - \varepsilon \leq \liminf_{n\to\infty} \frac{|\mathfrak{b}_\mu(x_{[0,n)})|}{n} \leq \limsup_{n\to\infty} \frac{|\mathfrak{b}_\mu(x_{[0,n)})|}{n} \leq \mathcal{H}(\mu) + \varepsilon$$

Pro $x \in \bigcap_{k>0} V(1/k)$ je tedy $\lim_{n\to\infty} |\mathfrak{b}_\mu(x_{[0,n)})|/n = \mathcal{H}(\mu)$ a tato množina má míru 1. □

Věta 106 *Pro každé prosté zobrazení $f : A^* \to B^*$ a každý ergodický proces platí*

$$\liminf_{n\to\infty} \frac{|f(X_{[0,n)})|}{n} \geq \mathcal{H}(X) \textsf{ s.j.}$$

Důkaz: Položme Nechť $\mathfrak{b}_B : B^* \to B^*$ je prefixový kód sestrojený v Tvrzení 68. Pro každé ε existuje n_ε takové že pro všechna $n > n_\varepsilon$ platí $|\mathfrak{b}_B(u)| \leq (1+\varepsilon)|u|$. Položme $g(u) = \mathfrak{b}_B(f(u))$, takže $g : A^* \to B^*$ je prefixový kód. Položme

$$V_n = \{u \in A^n : \ |g(u)| \leq -\log \mu(u) - 2\log n\}, \ \ V = \limsup_{n\to\infty} [V_n]$$

Pro g platí Kraftova nerovnost, tedy speciálně $\sum_{u\in A^n} 2^{-|g(u)|} \leq 1$, takže

$$\mu(V_n) \ \ \leq \ \sum_{u\in V_n} 2^{-|g(u)|} 2^{-2\log n} \leq \frac{1}{n^2}$$

Je tedy $\sum_{n>0} \mu(V_n) < \infty$ a $\mu(V) = 0$ podle Borel-Cantelliho lemma. Nechť $W \subseteq A^{\mathbb{N}}$ je množina sekvencí, pro které platí Entropická věta. Pak $\mu(W \setminus V) = 1$ a pro $x \in W \setminus V$ platí

$$\liminf_{n\to\infty} \frac{|f(x_{[0,n)})|}{n} \geq \liminf_{n\to\infty} \frac{|g(x_{[0,n)})|}{n(1+\varepsilon)} \geq \liminf_{n\to\infty} \frac{-\log \mu(x_{[0,n)}) - 2\log n}{n(1+\varepsilon)} \geq \frac{\mathcal{H}(X)}{1+\varepsilon}$$

takže $\liminf_{n\to\infty} |f(x_{[0,n)})|/n \geq \mathcal{H}(X)$. $\square$

Věta 107 *Existuje univerzální frekvenční prefixový kód* $\mathfrak{f} : A^* \to B^*$ *takový že pro každý ergodický proces platí*

$$\lim_{n\to\infty} \frac{|\mathfrak{f}(X_{[0,n)})|}{n} = \mathcal{H}(X) \text{ s.j.}$$

Důkaz: Pro danou abecedu A uvažujme funkce

$$k_n = \lfloor \log_{|A|}(\sqrt{n} - 1) \rfloor = \left\lfloor \frac{\log(\sqrt{n}-1)}{\log |A|} \right\rfloor, \ \ m_n = \lceil n/k_n \rceil.$$

Pro $k = k_n$, $m = m_n$, $u \in A^n$, $v \in A^k$ položme

$$|u|_v = |\{i < m : \ (uu)_{[ik,ik+k)} = v\}|, \ \ \mathfrak{P}_u^k(v) = |u|_v/m$$

Dále označme $\mathcal{C}^k(u) = \{w \in A^{|u|} : \ \mathfrak{P}_w^k = \mathfrak{P}_u^k\}$, $\ell(u) = \{w \in \mathcal{C}_u^k : \ w < u\}$. Pak platí $\ell(u) < |\mathcal{C}^k(u)| \leq 2^{m\cdot\mathcal{H}(\mathfrak{P}_u^k)} \leq |A|^{mk}$. Pro dané $u \in A^n$, položme

$$f(u) = \mathfrak{b}_0(|u|)(\mathfrak{b}_0(|u|_v))_{v\in A^k} \mathfrak{b}_{\mathbb{N}}(\ell(u))$$

Protože $|A|^k \leq \sqrt{n} - 1$, je

$$\begin{aligned} |f(u)| \ &\leq \ 2\log(|u|+1) + 1 + |A^k| \cdot (2\log(m+1)+1) \\ &\quad + m \cdot \mathcal{H}(\mathfrak{P}_u^k) + 2\log(m \cdot k \log |A| + 1) + 1 \\ &\leq \ (2\log(n+1)+1) \cdot \sqrt{n} + m \cdot \mathcal{H}(\mathfrak{P}_u^k) + 2\log m \cdot k + 2\log\log |A| + 3 \end{aligned}$$

Protože $\lim_{n\to\infty} \sqrt{n} \cdot \log(n+1)/n = 0$ a $\lim_{n\to\infty} m_n/nk_n = 1$, je

$$\limsup_{n\to\infty} \frac{|f(x_{[0,n)})|}{n} \leq \limsup_{n\to\infty} \frac{\mathcal{H}(\mathfrak{P}_{X_{[0,n)}}^{k_n})}{k_n}$$

Pro dané $\varepsilon > 0$ existuje k_ε takové že $\mathcal{H}(X_{[0,k)}) < k(\mathcal{H}(mu)+\varepsilon)$. Pro skoro všechna x existuje $N(x,\varepsilon)$ takové že pro každé $n > N(x,\varepsilon)$ je $\mathcal{H}(\mathfrak{P}_{X_{[0,n)}}^k) < n(\mathcal{H}(\mu)+\varepsilon)$. Pro dosti velká n je $k_n > k_\varepsilon$ takže

$$\limsup_{n\to\infty} \frac{\mathcal{H}(\mathfrak{P}_{X_{[0,n)}}^{k_n})}{k_n} \leq \mathcal{H}(\mu) + \varepsilon$$

a tedy $\limsup_{n\to\infty} |f(x_{[0,n)})|/n \leq \mathcal{H}(\mu)$ s.j. $\square$

5.5 Rekurenční kód

Posloupnost slov $y = (y_i)_{i<\ell(y)}$ je rozklad slova $u \in A^*$ jestliže u je konkatenace $u = y_0 \cdots y_{\ell(y)-1}$ a $y_i \neq y_j$ pro $i \neq j$. Označme $R(u)$ množinu všech rozkladů slova u.

Lemma 108 *Pro $a > 1$ platí*

$$m_k(a) = \sum_{j=1}^{k} a^j = \frac{a^{k+1} - a}{a - 1}, \quad n_k(a) = \sum_{j=1}^{k} j \cdot a^j = \frac{k \cdot a^{k+2} - (k+1)a^{k+1} + a}{(a-1)^2}$$

$$a^k < m_k(a) < \frac{a^{k+1}}{a-1}, \quad k \cdot a^k < n_k(a) < \frac{k \cdot a^{k+1}}{a-1}$$

Důkaz: $m_k(a)$ je součet geometrické posloupnosti. Pro $n_k(a)$ platí odhad $n_k(a) \leq \sum_{j=1}^{k} k \cdot a^j \leq \frac{k \cdot a^{k+1}}{a-1}$ a přesný vzorec

$$\begin{aligned} n_k(a) &= \sum_{j=1}^{k} a^j + \sum_{j=2}^{k} a^j + \cdots + \sum_{j=k}^{k} a^j = a \cdot \frac{a^k - 1}{a-1} + \cdots + a^k \cdot \frac{a-1}{a-1} \\ &= \frac{1}{a-1} \left(k \cdot a^{k+1} - \frac{a^{k+1} - a}{a-1} \right) = \frac{k \cdot a^{k+2} - (k+1)a^{k+1} + a}{(a-1)^2} \quad \square \end{aligned}$$

Lemma 109 $\forall \delta > 0, \forall k > 0, \exists N_{k,\delta}, \forall n > N_{k,\delta}, \forall u \in A^n, \forall y \in R(u)$

$$\sum \{|y_i| : \ i < \ell(y), |y_i| \leq k\} < n\delta$$

Důkaz: Položme $a = |A|$, $N_{k,\delta} = k \cdot a^{k+1}/\delta(a-1)$. Pak

$$\sum \{|y_i| : \ i < \ell(y), |y_i| \leq k\} < \sum_{j=1}^{k} j \cdot a^j < \frac{ka^{k+1}}{a-1} < \delta \cdot N_{k,\delta} < n\delta \ \square$$

Tvrzení 110 *Nechť $L(n)$ je nejkratší délka rozkladu slova délky n (v abecedě A). Pak*

$$\liminf_{n \to \infty} \frac{n}{L(n) \cdot \log n} \geq \frac{|A| - 1}{|A|^2 \cdot \log |A|}$$

Důkaz: Nechť $y = (y_i)_{i<\ell(y)}$ je rozklad slova $x \in A^n$. V posloupnosti y nahraďme delší slova kratšími tak aby výsledná posloupnost y' byla opět rozkladem (nějakého jiného slova x'), přitom ale aby platilo že pokud $1 \leq |u| < y'_i$, pak $u = y'_j$ pro nějaké $j < \ell(y')$. Pak $\ell(y') = \ell(y)$ a $|x'| \leq |x|$. Existuje k takové že $m_k(|A|) \leq \ell(y') < m_{k+1}(|A|)$, a $n \geq |x'| \geq n_k(|A|)$. Protože $n/\log n$ je rostoucí,

$$\frac{n}{L(n) \cdot \log n} \quad \geq \quad \frac{k \cdot a^k}{(k \cdot \log a + \log k)a^{k+2}/(a-1)} = \frac{k(a-1)}{a^2(k \cdot \log a + \log k)}$$

Odtud již plyne tvrzení. $\square$

Lemma 111 *Nechť μ je ergodická míra, $h = \mathcal{H}(\mu)$ a $\varepsilon > 0$. Pak existuje $V \subseteq A^*$ taková že platí $|V \cap A^k| \leq 2^{k(h+\varepsilon)}$, a pro posloupnost množin*

$$W_n(\varepsilon) = \left\{ x \in A^{\mathbb{N}} : \ \forall y \in R(x_{[0,n)}), \sum \{|y_i| : \ i < \ell(y) \ \& \ y_i \in V\} \geq (1-\varepsilon)n \right\}$$

platí $\mu(\liminf_{n\to\infty} W_n(\varepsilon)) = 1$.

Důkaz: Existuje $\delta < \varepsilon/2$ takové že $\delta + \mathcal{H}(\delta, 1-\delta) + \delta \cdot \log|A| < \varepsilon$. Podle Entropické věty existuje $d > 0$ takové, že pro množinu $M = \{u \in A^d : \mu(u) \geq 2^{-d(h+\delta)}\}$ platí $\mu(M) > 1 - \delta^2/8$. Pro $K = \lceil 2d/\delta \rceil + 1$ podle Lemma 109 existuje N_K takové, že pro každé $n > N_K$, $u \in A^n$ a každý rozklad $(y_i)_{i<\ell(y)}$ slova u platí

$$\sum\{|y_i| : i < \ell(y) \,\&\, |y_i| < K\} < n\delta/2.$$

Označme $V_n = V_n(\delta)$ množinu všech slov, která jsou δ-pokrývána množinou M a $V = \bigcup_n V_n$. Protože $|M| \leq 2^{d(h+\varepsilon)}$, je podle Lemma 102 $|V_n| \leq 2^{n(h+\varepsilon)}$. Podle Ergodické věty existuje množina U míry 1 taková, že pro každé $x \in U$ existuje $N(x) \geq N_K$ takové, že pro všechna $n > N(x)$ platí

$$|\{i < n : \sigma^i(x) \in M\}| \geq n(1 - \delta^2/4).$$

Nechť $x \in U$, $n > N(x)$ a nechť $(y_i)_{i<\ell(y)}$ je rozklad slova $x_{[0,n)}$. Označme $y_i = x_{[q_i, q_{i+1})}$,

$$P \;=\; \{j < \ell(y) : |y_j| > K \,\&\, |\{i < |y_j| - d : \sigma^{q_j+i}(x) \in M\}| < (|y_j| - d)(1 - \tfrac{\delta}{2})\}$$

a $Q = [0, \ell) \setminus P$. Předpokládejme sporem že platí $\sum_{j\in P} |y_j| > n\delta$. Protože $|P| \leq n/K$, je

$$\begin{aligned}
n(1 - \tfrac{\delta^2}{4}) &\leq |\{i < n : T^i(x) \in M\}| \leq \sum_{j\in P}(|y_j| - d)(1 - \tfrac{\delta}{2}) + d \cdot |P| + \sum_{j\in Q} |y_j| \\
&\leq n - \sum_{j\in P}(|y_j| - d)\frac{\delta}{2} = n + d|P|\frac{\delta}{2} - \frac{\delta}{2}\sum_{j\in P}|y_j| < n + \frac{nd\delta}{2K} - \frac{n\delta^2}{2} \\
&< n + \frac{n\delta^2}{4} - \frac{n\delta^2}{2} = n - \frac{n\delta^2}{4}
\end{aligned}$$

a to je spor. Je tedy $\sum_{j\in P} |y_j| \leq n\delta$. Podle Lemma 102, pokud $y_j \notin V$, pak buď $y_j \in P$ nebo $|y_j| \leq 2d/\delta < K$. Odtud

$$\begin{aligned}
\sum\{|y_j| : j < \ell \,\&\, y_j \notin V\} &= \sum\{|y_j| : j < \ell \,\&\, |y_j| < K\} + \sum\{|y_j| : j \in P\} \\
&\leq n\delta + n\delta \leq n\varepsilon
\end{aligned}$$

takže $x \in W_n(\varepsilon)$. Ukázali jsme tedy $U \subseteq \liminf_{n\to\infty} W_n(\varepsilon)$ takže $\mu(\liminf_{n\to\infty} W_n(\varepsilon)) = 1$. □

Tvrzení 112 *Pro ergodickou míru μ s entropií $h = \mathcal{H}(\mu)$ položme*

$$U_n(\varepsilon) = \left\{x \in A^{\mathbb{N}} : \forall y \in R(x_{[0,n)}), \sum\{|y_i| : |y_i| < \tfrac{\log n}{h+\varepsilon}\} \leq \varepsilon \cdot n\right\}$$

Pak $\mu(\liminf_{n\to\infty} U_n(\varepsilon)) = 1$.

Důkaz: Pro $\delta > 0$ existuje podle Lemma 111 $V \subseteq A^*$ takové že $|V \cap A^k| \leq 2^{k(h+\delta)}$, a $W(\delta) = \liminf_{n\to\infty} W_n(\delta)$ má míru 1. Pro $x \in W(\delta)$ tedy existuje $N(x)$ takové že pro každé $n > N(x)$ je $x_{[0,n)} \in W_n(\delta)$. Nechť $y \in R(x_{[0,n)})$ takže $\sum\{|y_i| : y_i \in V\} > (1-\delta)n$. S použitím nerovnosti $\sum_{j\leq k} j \cdot a^j \leq k\frac{a}{a-1}a^k$, která platí pro $a > 1$ dostáváme

$$\begin{aligned}
&\sum\left\{|y_i| : i < \ell(y) \,\&\, |y_i| < \frac{(1-\delta)\log n}{h+\delta}\right\} \\
&\qquad\leq \sum\{|y_i| : i < \ell(y) \,\&\, y_i \notin V\} + \sum\left\{k \cdot |V \cap A^k| : k < \frac{(1-\delta)\log n}{h+\delta}\right\} \\
&\qquad\leq \delta \cdot n + \sum\left\{k \cdot 2^{(h+\delta)k} : k < \frac{(1-\delta)\log n}{h+\delta}\right\} \\
&\qquad\leq \delta \cdot n + \frac{2^{h+\delta}}{2^{h+\delta}-1} \cdot \frac{(1-\delta)\log n}{h+\delta} \cdot n^{1-\delta} = \varphi_\delta(n)
\end{aligned}$$

Pro dané ε zvolme nyní δ takové že $\frac{1}{h+\varepsilon} < \frac{1-\delta}{h+\delta}$ a pro všechna dosti velká n platí $\varphi_\delta(n) < \varepsilon \cdot n$. Pak $x \in U_n(\varepsilon)$, takže $W(\delta) \subseteq \liminf_{n\to\infty} U_n(\varepsilon)$, takže $\mu(\liminf_{n\to\infty} U_n(\varepsilon)) = 1$ □

Lemma 113 *Pro $y > 1/\varepsilon^2$ a $3 \le x \le \varepsilon \cdot y$ platí $\frac{x}{\log x} \le 2\varepsilon \cdot \frac{y}{\log y}$.*

Věta 114 *Pro rekurenční kód $\mathfrak{r}$ a každý ergodický proces X platí*

$$\lim_{n\to\infty} \frac{|\mathfrak{r}(X_{[0,n)})|}{n} = \mathcal{H}(X) \text{ s.j.}$$

Důkaz: Je-li $y = (y_i)_{i\le\ell}$ rozbor slova $X_{[0,n)}$, jsou slova y_i navzájem různá, s možnou vyjímkou posledního slova y_ℓ. Toto slovo se však shoduje s nějakým předcházejícím y_i, takže $|y_\ell| \le n/2$. Použijeme Tvrzení 112 na rozklad $y' = (y_i)_{i<\ell}$ délky $m \ge n/2$. Pro dané $\varepsilon > 0$ zvolme $\delta < \varepsilon$ takové že $\delta + 2\delta|A|^2 \log|A| < \varepsilon$. Nechť $x \in U(\delta) = \liminf_{n\to\infty} U_n(\delta)$ a $N(x)$ takové že $x \in U_n(\delta)$ pro všechna $n > N(x)$. Pro dané $\varepsilon > 0$ zvolme $\delta < \varepsilon$ takové že $\delta + 2\delta|A| \log|A| < \varepsilon$. Označme

$$L_0 = \left\{ i < \ell : \ |y_i| < \frac{\log m}{h+\delta} \right\}, \quad L_1 = \left\{ i < \ell : \ |y_i| \ge \frac{\log m}{h+\delta} \right\}$$

$M_0 = \sum_{i\in L_0} |y_i|$, $M_1 = \sum_{i\in L_1} |y_i|$. Pak $M_1 \ge |L_1| \log m/(h+\delta)$ a podle Tvrzení 112 pro $n > N(x)$ platí $M_0 \le \delta \cdot m$. Podle Tvrzení 110 je $M_0/|L_0| \log M_0 \ge 1/|A|^2 \log|A|$, takže

$$\begin{aligned} \ell &= |L_0| + |L_1| \le \frac{2M_0|A|^2 \log|A|}{\log M_0} + \frac{M_1(h+\delta)}{\log m} \\ &\le \frac{m \cdot 2\delta|A|^2 \log|A| + m(h+\delta)}{\log m} \le \frac{n(h+\varepsilon)}{\log n - 1} \end{aligned}$$

Pro délku rekurenčního kódu platí

$$|\mathfrak{r}(x_{[0,n)})| \le (\ell+1)(\log n + 1) \le n(h+\varepsilon) \cdot \frac{\log n + 1}{\log n - 1} + \log n + 1$$

takže $\limsup_{n\to\infty} |\mathfrak{r}(X_{[0,n)})|/n \le h + \varepsilon$ skoro jistě. □

5.6 Doba návratu

Definice 35 *Doba vstupu bodu $x \in A^{\mathbb{N}}$ do množiny $V \subseteq A^{\mathbb{N}}$ je*

$$R_V(x) = \inf\{k > 0 : \ \sigma^k(x) \in V\}$$

Doba návratu bodu $x \in A^{\mathbb{N}}$ do cylindru $[x_{[0,n)}]$ je

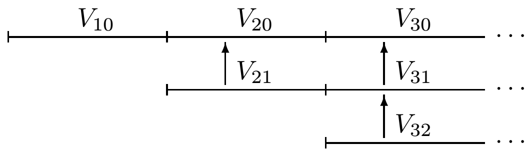
$$R_n(x) = \inf\{i > 0 : \ x_{[i,i+n)} = x_{[0,n)}\}$$

Infimum prázdné množiny definujeme jako ∞.

Věta 115 (Poincaré) *Je-li μ stacionární míra, a $\mu(V) > 0$, je doba návratu do V skoro jistě konečná*

$$\mu\{x \in V : \ R_V(x) = \infty\} = 0$$

Důkaz: Předpokládejme, že množina $V_\infty = \{x \in V : \ R_V(x) = \infty\}$ má kladnou míru. Množiny $\sigma^{-k}(V_\infty)$ jsou navzájem disjunktní. Pokud by totiž bylo $x \in \sigma^{-j}(V_\infty) \cap \sigma^{-k}(V_\infty)$ pro $j < k$, bylo by $\sigma^j(x) \in V_\infty \cap \sigma^{-(k-j)}(V_\infty)$ a to je spor. Protože množiny $\sigma^{-k}(V_\infty)$ mají všechny stejnou míru, musí být tato míra nulová. □



Obrázek 20: Průměrná doba návratu

Věta 116 (Kac) *Nechť μ je ergodická míra, $\mu(V) > 0$, a nechť $\mu_V(U) = \mu(U)/\mu(V)$ je podmíněná míra na V. Pak*

$$\mathbb{E}_\mu(R_V \cdot \chi_V) = \int_V R_V(x)\,d\mu = 1, \quad \mathbb{E}_{\mu_V}(R_V) = \int_V R_V\,d\mu_V = \frac{1}{\mu(V)}$$

Důkaz: Položme $V_1 = \sigma^{-1}(V) \cap V = \{x \in V : \tau_V(x) = 1\}$,

$$V_n = (\sigma^{-n}(V) \cap V) \setminus \bigcup_{k<n} \sigma^{-k}(V) = \{x \in V : \tau_V(x) = n\}$$

Množiny V_n jsou navzájem disjunktní a podle Věty 115 je $\mu(V_\infty) = 0$, takže $\sum_{n=1}^\infty \mu(V_n) = \mu(V)$. Pro $0 \leq k < n$ položme $V_{nk} = \sigma^{-k}(V_n)$. Pak množiny $(V_{nk})_{0\leq k<n}$ jsou navzájem disjunktní (obrázek 20) a jejich sjednocení je $\bigcup_{n\geq 0} \sigma^{-n}(V \setminus V_\infty)$ a $\sigma(V) \subseteq V_{10} \cup V_{21} \cup V_{32} \cup \cdots$. Protože μ je ergodická, je

$$1 \;=\; \mu\left(\bigcup_{n\geq 0} \sigma^{-n}(V)\right) \geq \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} \mu(V_{nk}) = \sum_{n=1}^{\infty} n \cdot \mu(V_n) = \int_V \tau_V(x)\,d\mu. \;\square$$

Je-li μ ergodická míra, $u \in A^n$ a $\mu(u) > 0$, je

$$\int R_n(x)\,d\mu \;=\; \sum \left\{ \int_{[u]} R_n\,d\mu : \; u \in A^n, \mu(u) > 0 \right\} = |\{u \in A^n : \; \mu(u) > 0\}|$$

Věta 117 (Ornstein-Weiss) *Je-li X ergodický proces, je*

$$\lim_{n\to\infty} \frac{\log R_n(x)}{n} = \mathcal{H}(X) \textsf{ s.j.}$$

Důkaz: Pro $x \in A^{\mathbb{N}}$ definujme funkce

$$\underline{R}(x) = \liminf_{n\to\infty} \frac{\log R_n(x)}{n}, \quad \overline{R}(x) = \limsup_{n\to\infty} \frac{\log R_n(x)}{n},$$

Protože $R_{n-1}(\sigma(x)) \leq R_n(x)$, je $\underline{R}(\sigma(x)) \leq \underline{R}(x)$, $\overline{R}(\sigma(x)) \leq \overline{R}(x)$, takže tyto funkce jsou skoro všude konstantní a existují $\underline{r} \leq \overline{r}$, takové že

$$\underline{R}(x) = \underline{r} \textsf{ s.j.}\,, \quad \overline{R}(x) = \overline{r} \textsf{ s.j.}$$

Ukážeme $\overline{r} \leq h = \mathcal{H}(X)$. Pro $\varepsilon > 0$ položme

$$\begin{aligned} D_n &= \{x \in A^{\mathbb{N}} : \; R_n(x) > 2^{n(h+\varepsilon)}\}, \quad D_n(u) = D_n \cap [u] \\ T_n &= \{u \in A^n : \; \mu(u) \geq 2^{-n(h+\varepsilon/2)}\} \end{aligned}$$

Podle Entropické věty je $\mu(\liminf_{n\to\infty}[T_n]) = 1$. Pro $u \in A^n$ jsou množiny $\sigma^{-k}(D_n(u))$ pro $k < 2^{n(h+\varepsilon)}$ navzájem disjunktní a mají stejnou míru, takže $\mu(D_n(u)) \leq 2^{-n(h+\varepsilon)}$. Dále platí

$$\mu(D_n \cap [T_n]) \leq \sum_{u\in T_n} \mu(D_n(u)) \leq |T_n| \cdot 2^{-n(h+\varepsilon)} \leq 2^{-n\varepsilon/2}$$

protože $|T_n| \leq 2^{n(h+\varepsilon/2)}$. Podle Borel-Cantelliho lemma je $\mu(\limsup_{n\to\infty}(D_n \cap [T_n])) = 0$, takže $\mu(\limsup_{n\to\infty} D_n) = 0$. To znamená $\overline{r} \leq h + \varepsilon$ a tedy $\overline{r} \leq h$.
Ukážeme $h \leq \underline{r}$ pro $\varepsilon > 0$ zvolme δ takové že $\delta + \mathcal{H}(\delta, 1-\delta) + \delta \log |A| < \varepsilon$. Položme

$$V_L = \left\{ x \in A^{\mathbb{N}} :\ \exists j < L, \frac{\log R_j(x)}{j} < \underline{r} + \delta \right\}$$

Existuje $L > 0$ takové že $\mu(V_L) > 1 - \delta/4$. Položme $L_1 = L + 2^{L(\underline{r}+\delta)}$ a

$$W_n = \left\{ x \in A^{\mathbb{N}} :\ |\{i < n :\ \sigma^i(x) \in V_L\}| > (1 - \tfrac{\delta}{2})n \right\},\ \ W = \liminf_{n\to\infty} W_n$$

Pak $\mu(W) = 1$. Pro $x \in W$ existuje $N(x)$ takové že pro $n > N(x)$ je $x \in W_{n-L_1}$. Položme

$$M = \{u \in A^* :\ \exists j < L, |u| < 2^{j(\underline{r}+\delta)} + j, u_{[0,j)} = u_{[|u|-j,|u|)}\}$$

Pak $|M \cap A^n| \leq 2^{n(\underline{r}+\delta)}$. Pro množinu $D_n(\delta)$ všech slov délky n která jsou δ-pokryta M podle Lemma 102 platí

$$|D_n(\delta)| \leq |A|^{n\delta} \cdot 2^{n\mathcal{H}(\delta)} \cdot 2^{n(\underline{r}+\delta)} \leq 2^{n(\underline{r}+\varepsilon)}.$$

Pokud $T_i(x) \in V_L$, pak existuje $j < L$ a $k < 2^{j(\underline{r}+\delta)} + j \leq L_1$ takové že $x_{[i,i+j)} = x_{[i+k-j,i+k)}$, takže $x_{[i,i+k)} \in M$. Pokud $x \in W_{n-L_1}$, pak

$$|\{i < n - L_1 :\ \exists k < L_1, x_{[i,i+k)} \in M\}| > (1 - \tfrac{\delta}{2})(n - L_1)$$

a podle Lemma 102 je $x_{[0,n)} \in D_n(\delta)$. Je tedy $W_{n-L_1} \subseteq D_n(\delta)$, takže

$$\mu(\liminf_{n\to\infty}[D_n(\delta)]) = 1.$$

Pro množiny

$$U_n = \{u \in A^n :\ \mu(u) < 2^{-n(h-\varepsilon)}\}$$

platí $\mu(\liminf_{n\to\infty}[U_n]) = 1$ a tedy $\mu(\liminf_{n\to\infty}[U_n \cap D_n(\delta)]) = 1$. Protože

$$\mu(U_n \cap D_n) \leq |D_n(\delta)| \cdot 2^{-n(h-\varepsilon)} \leq 2^{n(\underline{r}-h+2\varepsilon)},$$

je $\underline{r} - h + 2\varepsilon \geq 0$. Protože to platí pro každé ε, je $h \leq \underline{r}$. $\square$

6 Symbolická dynamika

V některých případech máme o textu který chceme přenášet, nějaké strukturální informace. Jedná-li se o text v nějakém umělém jazyce daném formální gramatikou, víme že některá slova se v něm nemohou vyskytnout. Množiny (nekonečných) slov, ve kterých se nevyskytují nějaká konečná slova, se nazývají posuny.

Definice 36 *Posun nad abecedou A je každá neprázdná množina tvaru*

$$\Sigma_F = \{x \in A^{\mathbb{N}} : \forall u \sqsubseteq x, u \notin F\}$$

kde $F \subseteq A^*$ je nějaká množina zakázaných slov. Jazyk posunu $\Sigma \subseteq A^{\mathbb{N}}$ je

$$\mathcal{L}(\Sigma) = \bigcup_{n\geq 0} \mathcal{L}^n(\Sigma), \quad kde \;\; \mathcal{L}^n(\Sigma) = \{u \in A^n : \exists x \in \Sigma, u \sqsubseteq x\}$$

Funkce složitosti posunu Σ je $P_\Sigma(n) = |\mathcal{L}^n(\Sigma)|$

Posun zlatého řezu je $\Sigma_{\{11\}}$ a $\mathcal{L}(\Sigma_{\{11\}}) = \{\lambda, 0, 1, 00, 01, 10, 000, 001, 010, 100, 101, \ldots\}$. Funkce složitosti tvoří Fibonacciovu posloupnost $P(0) = 1$, $P(1) = 2$, $P(2) = 3$, $P(3) = 5$. Při zápisu dat na magnetické disky se klade podmínka aby nulové bloky nebyly ani příliš krátké ani příliš dlouhé. Pro dané $1 \leq k \leq l$ definujme posun

$$\Sigma_{k,l} = \Sigma_F, \text{ kde } F = \{11, 101, 1001, \ldots, 10^{k-1}1, 0^{l+1}\}$$

Sudý posun má zakázaná slova $F = \{01^{2n+1}0, : \; n \geq 0\}$. Mezi dvěma po sobě jdoucími výskyty nuly může být jen sudý počet jedniček. Funkce složitosti je $P(0) = 1$, $P(1) = 2$, $P(2) = 4$, $P(3) = 7$, $\mathcal{L}(\Sigma) = \{\lambda, 0, 1, 00, 01, 10, 11, 000, 001, 011, 100, 101, 110, 111, \ldots\}$. Bezkontextový posun abecedě $A = \{0, 1, 2\}$ má zakázaná slova $F = \{01^i2^j0 : \; i \neq j\}$.

Definice 37

(1) Posun Σ je konečného typu, existuje-li konečná množina $F \subset A^*$, pro kterou $\Sigma = \Sigma_F$.

(2) Posun konečného typu je řádu k, pokud existuje $F \subseteq A^{k+1}$ taková že $\Sigma = \Sigma_F$.

(3) Markovský posun je posun řádu 1.

(4) Posun Σ je sofický, pokud $\mathcal{L}(\Sigma)$ je regulární jazyk.

Posun zlatého řezu je řádu 1, Posun $\Sigma_{k,l}$ je řádu l. Každý posun konečného typu je sofický. Sudý posun není konečného typu, ale je sofický. Bezkontextový posun není sofický.

Věta 118 *Je-li $P : \mathbb{N} \to \mathbb{N}$ funkce složitosti posunu Σ, existuje limita*

$$\mathcal{H}(\Sigma) = \lim_{n\to\infty} \frac{\log P(n)}{n}$$

Tato limita se nazývá entropie posunu Σ.

Důkaz: Je-li $u \in A^n$, $v \in A^m$ a $uv \in \mathcal{L}(\Sigma)$, pak u i v náleží do $\mathcal{L}(\Sigma)$. Z toho plyne $P(n+m) \leq P(n) \cdot P(m)$. Pro $a_n = \log P(n)$ tedy platí $a_{n+m} \leq a_n + a_m$. Pro pevné $m > 0$ a $n > 0$ položme $q_n = \lceil n/m \rceil$, takže $(q_n - 1) \cdot m < n \leq q_n \cdot m$ a $a_n \leq q_n \cdot a_m$. Odtud

$$\begin{aligned} \limsup_{n\to\infty} \frac{a_n}{n} &\leq \limsup_{n\to\infty} \frac{q_n}{q_n - 1} \cdot \frac{a_m}{m} = \frac{a_m}{m} \\ \limsup_{n\to\infty} \frac{a_n}{n} &\leq \liminf_{n\to\infty} \frac{a_m}{m} \;\; \square \end{aligned}$$

6.1 Prostor symbolických měr

Na množině $A^{\mathbb{N}}$ definujeme vzdálenost

$$d(x,y) = 2^{-n} \text{ kde } n = \min\{i \geq 0 : \; x_i \neq y_i\}.$$

Cantorova množina C je podmnožina reálné přímky kterou získáme tak, že z jednotkového intervalu $\mathbb{I} = [0,1]$ odstraníme prostřední otevřený interval $(\frac{1}{3}, \frac{2}{3})$ a tento postup opakujeme s intervaly které zbydou (obrázek 21).

$$C = [0,1] \setminus (\tfrac{1}{3}, \tfrac{2}{3}) \setminus (\tfrac{1}{9}, \tfrac{2}{9}) \setminus (\tfrac{7}{9}, \tfrac{8}{9}) \ldots$$

Cantorova množina sestává z čísel intervalu $\mathbb{I}$, které lze vyjádřit v trojkové soustavě pouze s použitím číslic $0, 2$, tj. ve tvaru $\sum_{i\geq 0} a_i 3^{-i-1}$, kde $a_i \in \{0,2\}$.

Obrázek 21: Cantorova množina

Tvrzení 119 *Každý symbolický prostor $A^{\mathbb{N}}$ je homeomorfní Cantorově množině.*

Důkaz: Pro binární abecedu B definujeme vzájemně jednoznačné zobrazení $\varphi : B^{\mathbb{N}} \to C$ předpisem $\varphi(x) = \sum_{i=0}^{\infty} 2x_i 3^{-i-1}$. Pokud $d(x,y) = 2^{-n}$, pak $x_i = y_i$ pro $i < n$, $x_n \neq y_n$, a

$$\begin{aligned}
|\varphi(x) - \varphi(y)| &= \left| \sum_{i=n}^{\infty} 2(x_i - y_i) 3^{-i-1} \right| \leq 2 \sum_{i=n}^{\infty} 3^{-i-1} = \frac{2 \cdot 3^{-n-1}}{1 - \frac{1}{3}} = 3^{-n} \\
|\varphi(x) - \varphi(y)| &\geq 2 \cdot |x_n - y_n| \cdot 3^{-n-1} - 2 \left| \sum_{i=n+1}^{\infty} (x_i - y_i) \right| \\
&\geq 2 \cdot 3^{-n-1} - 2 \sum_{i=n+1}^{\infty} 3^{-i-1} = 3^{-n-1}
\end{aligned}$$

Dostáváme tedy

$$\begin{aligned}
d(x,y) \leq 2^{-n} &\Rightarrow |\varphi(x) - \varphi(y)| \leq 3^{-n} \\
|\varphi(x) - \varphi(y)| < 3^{-n-1} &\Rightarrow d(x,y) < 2^{-n}
\end{aligned}$$

To znamená že jak φ tak φ^{-1} jsou spojité. Pro obecnou abecedu $A = \{a_0, \ldots, a_{k-1}\}$, kde $k \geq 3$ zvolme libovolný úplný prefixový kód $f : A \to B^+$, například $f(a_i) = 0^i 1$. Pak jeho rozšíření $f : A^{\mathbb{N}} \to B^{\mathbb{N}}$ je homeomorfismus. Platí totiž $d(f(x), f(y)) \leq d(x,y)$ a pokud $d(f(x), f(y)) < 2^{-nk}$, pak $d(x,y) < 2^{-n}$. $\square$

Symbolický prostor je kompaktní, protože Cantorova množna je uzavřená a omezená. Kompaktnost prstoru $A^{\mathbb{N}}$ také plyne z Tichonovy věty. Diskrétní prostor A je kompaktní a

$A^{\mathbb{N}}$ je jeho spočetný produkt. Cylindr $[u] = \{x \in A^{\mathbb{N}} : x_{[0,n)} = u\}$ slova $u \in A^*$ je obojetná množina (tj. otevřená a uzavřená a platí

$$x \in [u] \implies [u] = B_{2^{-n+1}}(x) = \{y \in \mathbf{2}^{\mathbb{N}} : d(y,x) < 2^{-n+1}\}$$

Posun $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ definované předpisem $\sigma(x)_i = x_{i+1}$ je spojité zobrazení a platí $d(\sigma(x), \sigma(y)) \leq d(x, y)$.

Tvrzení 120 *Neprázdná množina $\Sigma \subseteq A^{\mathbb{N}}$ je posun právě když je uzavřená a σ-invariantní, tj. $\sigma(\Sigma) \subseteq \Sigma$.*

Důkaz: Nechť $F \subseteq A^*$ a nechť Σ_F je neprázdná. Zřejmě platí $\sigma(\Sigma_F) \subseteq \Sigma_F$. Jestliže $x \in A^{\mathbb{N}} \setminus \Sigma_F$, pak $x_{[n,m)} \in F$ pro nějaké $0 \leq n < m$. Pak $[x_{[0,m)}] \subseteq A^{\mathbb{N}} \setminus \Sigma_F$, takže $A^{\mathbb{N}} \setminus \Sigma_F$ je otevřená a tedy Σ_F je uzavřená.
Naopak předpokládejme že Σ je uzavřená a σ-invariantní a položme

$$F = \{u \in A^* : \forall x \in \Sigma, u \not\sqsubseteq x\}$$

Je-li $x \in \Sigma$, pak žádné $u \sqsubseteq x$ nenáleží do F, takže $x \in \Sigma_F$. Ukázali jsme tedy $\Sigma \subseteq \Sigma_F$. Předpokládejme sporem že existuje $x \in \Sigma_F \setminus \Sigma$. Protože $A^{\mathbb{N}} \setminus \Sigma$ je otevřená množina, existuje n takové že $[x_{[0,n)}] \subseteq A^{\mathbb{N}} \setminus \Sigma$. Ukážeme že $x_{[0,n)}$ náleží do F. Kdyby do něj nenáleželo, existovalo by $y \in \Sigma$ takové že $y_{[k,k+n)} = x_{[0,n)}$ pro nějaké $k \geq 0$. Protože Σ je σ-invariantní, je $\sigma^k(y) \in \Sigma$, ale $\sigma^k(y) \subseteq [x_{[0,n)}] \subseteq A^{\mathbb{N}} \setminus \Sigma$ a to je spor. $\square$

Připomeňme že symbolická míra je zobrazení $\mu : A^* \to [0,1]$ které splňuje podmínky kompatibility.

Definice 38 *Vzdálenost dvou symbolických měr definujeme jako*

$$d(\mu, \nu) = \sum_{n=0}^{\infty} \max\{|\mu(u) - \nu(u)| : u \in A^n\} \cdot 2^{-n-1}$$

Prostor symbolických měr nad abecedou A značíme $\mathcal{M}(A^{\mathbb{Z}})$. Podprostor stacionárních měr značíme $\mathcal{M}_\sigma(A^{\mathbb{Z}})$.

Pro posloupnost stacionárních měr platí $\lim_{n\to\infty} \mu_n = \mu$ právě když $\lim_{n\to\infty} \mu_n(u) = \mu(u)$ pro každé $u \in A^*$. Prostor symbolických měr je podprostor Hilbertovy krychle $[0,1]^{A^*}$, tj. spočetného součinu jednotkových intervalů. Podle Tichonovovy věty je Hilbertova krychle kompaktní prostor. Prostory $\mathcal{M}(A^{\mathbb{Z}})$ a $\mathcal{M}_\sigma(A^{\mathbb{Z}})$ jsou její uzavřené podprostory, takže jsou také kompaktní. Pro $x \in A^{\mathbb{N}}$ definujeme bodovou míru $\delta_x \in \mathcal{M}(A^{\mathbb{Z}})$ předpisem

$$\delta_x(u) = \begin{cases} 0 & \text{pro} \quad u \not\sqsubseteq x \\ 1 & \text{pro} \quad u \sqsubseteq x \end{cases}$$

Nechť $x, y \in A^{\mathbb{N}}$ jsou různé body a $m = \min\{i \geq 0 : x_i \neq y_i\}$. Pak

$$d(\delta_x, \delta_y) \quad = \quad \sum_{n=m}^{\infty} 2^{-m-1} = 2^{-n} = d(x, y)$$

Symbolický prostor $A^{\mathbb{N}}$ je tedy vnořen do prostoru $\mathcal{M}(A^{\mathbb{Z}})$. Je-li $\mu \in \mathcal{M}(A^{\mathbb{Z}})$, je míra $\sigma\mu$ definována předpisem

$$(\sigma\mu)(u) = \sum_{a \in A} \mu(au)$$

Míra $\mu \in \mathcal{M}(A^{\mathbb{Z}})$ je tedy stacionární právě když $\sigma\mu = \mu$.

6.2 Variační princip

Definice 39 *Nosič Σ_μ symbolické míry μ nad abecedou A je posun, jehož zakázaná slova jsou slova s nulovou pravděpodobností.*

$$\Sigma_\mu = \{x \in A^{\mathbb{N}} : \ \forall u \sqsubseteq x, \mu(u) > 0\}$$

Tvrzení 121 *Je-li μ stacionární míra, je $\mathcal{H}(\mu) \le \mathcal{H}(\Sigma_\mu)$.*

Důkaz: Pro $u \notin \mathcal{L}^n(\Sigma_\mu)$ je $\mu(u) = 0$, takže $(\mu(u))_{u \in A^n}$ je pravděpodobnostní rozložení na množině $\mathcal{L}^n(\Sigma_\mu)$. Odtud $\mathcal{H}(\mu|A^n) \le \log P(n)$, takže $\mathcal{H}(\mu) \le \mathcal{H}(\Sigma_\mu)$. $\square$

Věta 122 (Variační princip) *Pro každý posun Σ existuje stacionární míra μ taková že $\Sigma_\mu \subseteq \Sigma$ a $\mathcal{H}(\mu) = \mathcal{H}(\Sigma)$.*

Důkaz: Zvolme libovolný bod $z \in A^{\mathbb{N}}$ a sestrojme míry ν_n, μ_n předpisem

$$\nu_n = \frac{1}{|\mathcal{L}^n(\Sigma)|} \sum_{u \in \mathcal{L}^n(\Sigma)} \delta_{uz}, \quad \mu_n = \frac{1}{n} \sum_{k<n} \sigma^k \nu_n$$

Pro každé $u \in \mathcal{L}^n(\Sigma)$ tedy platí $\nu_n(u) = 1/|\mathcal{L}^n(\Sigma)|$. Protože prostor $\mathcal{M}(A^{\mathbb{Z}})$ je kompaktní, existuje limita vybrané posloupnosti $\mu = \lim_{i\to\infty} \mu_{n_i}$. Ukážeme že μ je stacionární a $\mathcal{H}(\mu) = \mathcal{H}(\Sigma)$. Pro každé $u \in A^*$ platí

$$\lim_{n\to\infty} |\sigma\mu_n(u) - \mu_n(u)| = \lim_{n\to\infty} |\sigma^n \nu_n(u) - \nu_n(u)|/n = 0$$

takže $\sigma\mu = \mu$ a $\mu \in \mathcal{M}_\sigma(A^{\mathbb{Z}})$. Vypočtěme nyní entropii $\mathcal{H}(\mu)$. Pro pevné m a $n > m$ položme $n = qm + r$ kde $0 \le r < m$. Nechť $X_i : A^{\mathbb{N}} \to A$ je projekce $X_i(x) = x_i$. Pro každou míru $\nu \in \mathcal{M}(A^{\mathbb{Z}})$ je X náhodný proces na $(A^{\mathbb{N}}, \mathcal{B}, \nu)$ s rozdělením ν. Z Tvrzení 7 plyne

$$\begin{aligned}
\mathcal{H}_{\mu_n}(X_{[0,m)}) &\ge \frac{1}{n}\sum_{k<n} \mathcal{H}_{\nu_n}(\sigma^k X_{[0,m)}) \ge \frac{1}{n}\sum_{i<m}\sum_{j<q} \mathcal{H}_{\nu_n}(T^{jm+i}X_{[0,m)}) \\
&\ge \frac{1}{n}\sum_{i<m} \mathcal{H}_{\nu_n}(X_{[i,qm+i)}) \ge \frac{1}{n}\sum_{i<m}(\mathcal{H}_{\nu_n}(X_{[0,n)}) - 2m\log|A|) \\
&= \frac{1}{n}\sum_{i<m}(\log P(n) - 2m\log|A|) = \frac{m}{n}\log P(n) - \frac{2m^2}{n}\log|A|
\end{aligned}$$

Odtud

$$\begin{aligned}
\mathcal{H}_\mu(X_{[0,m)}) &= \lim_{i\to\infty} \mathcal{H}_{\mu_{n_i}}(X_{[0,m)}) \ \ge \ m\mathcal{H}(\Sigma) \\
\mathcal{H}_\mu(X) &= \lim_{m\to\infty} \frac{\mathcal{H}_\mu(X_{[0,m)})}{m} \ \ge \ \mathcal{H}(\Sigma)
\end{aligned}$$

Opačná nerovnost plyne z Tvrzení 121. $\square$

6.3 Okénkové kódy

Definice 40 *Nechť $\Sigma \subseteq A^{\mathbb{N}}$, $\Theta \subseteq B^{\mathbb{N}}$ jsou posuny. Okénkový kód je spojité zobrazení $F : \Sigma \to \Theta$, které komutuje se zobrazím posunu, tj. $\sigma_B F = F\sigma_A$. Říkáme že posuny Σ a Θ jsou konjugované, existuje-li mezi nimi bijektivní okénkový kód.*

Okénkovým kódům $F : A^{\mathbb{N}} \to A^{\mathbb{N}}$ se též říká celulární automaty.

Věta 123 (Hedlund) *$F : \Sigma \to \Theta$ je okénkový kód právě když existuje $r \geq 0$ a zobrazení (lokální pravidlo) $f : \mathcal{L}^{r+1}(\Sigma) \to \mathcal{L}^1(\Theta)$ takové že $F(x)_i = f(x_{[i,i+r]})$.*

Důkaz: Jestliže F je spojité, pak pro $\varepsilon = 1$ existuje $\delta = 2^{-r}$ takové že pro $d(x, y) \leq 2^{-r}$ je $d(F(x), F(y)) < 1$, tj. $F(x)_0 = F(y)_0$. To znamená že existuje $f : \mathcal{L}^{r+1}(\Sigma) \to \mathcal{L}^1(\Theta)$ takové že $F(x)_0 = f(x_{[0,r]})$. Protože F komutuje se zobrazením posunu, je

$$F(x)_i(x) = \sigma^i(F(x))_0 = F(\sigma^i(x))_0 = f(\sigma^i(x)_{[0,r]}) = f(x_{[i,i+r]}$$

Naopak je-li $F(x)_i = f(x_{[i,i+r]})$, je F spojité a komutuje se zobrazením posunu. □

Okénkový kód $\Phi_n : A^{\mathbb{N}} \to (A^n)^{\mathbb{N}}$ definovaný předpisem

$$\Phi_n(x)_i = x_{[i,i+n)}$$

je bijektivní. Je-li $\Sigma \subseteq A^{\mathbb{N}}$ posun, je $\Sigma^{(n)} = \Phi_n(\Sigma) \subseteq (A^n)^{\mathbb{N}}$ konjugovaný posun. Spojitý obraz kompaktní množiny je totiž kompaktní a tedy uzavřená množina.

Tvrzení 124 *Je-li $F : \Sigma \to \Theta$ prostý okénkový kód, je $\mathcal{H}(\Sigma) \leq \mathcal{H}(\Theta)$. Je-li $F : \Sigma \to \Theta$ surjektivní okénkový kód, je $\mathcal{H}(\Sigma) \geq \mathcal{H}(\Theta)$.*

Důkaz: Nechť $F(x)_i = f(x_{[i,i+r]})$. Je-li F prosté a $x_{[0,n)} \neq y_{[0,n)}$, pak $F(x)_{[0,n)} \neq F(y)_{[0,n)}$, takže $P_\Sigma(n) \leq P_\Theta(n)$ a $\mathcal{H}(\Sigma) \leq \mathcal{H}(\Theta)$. Je-li F surjektivní, pak pro každé $u \in \mathcal{L}^n(\Theta)$ existuje $v \in f^{-1}(u)$, takže $P_\Sigma(n + r) \geq P_\Theta(n)$ a $\mathcal{H}(\Sigma) \geq \mathcal{H}(\Theta)$. □

6.4 Markovské posuny

Každý posun konečného typu je konjugovaný Markovskému posunu (řádu 1). Je-li totiž $\Sigma \subseteq A^{\mathbb{N}}$ posun řádu k, je $\Sigma^{(k)}$ posun řádu 1 v abecedě A^k. Markovský posun je dán binární přechodovou maticí M nad A, kde

$$M(a, b) = \begin{cases} 0 & \text{pro } ab \notin \mathcal{L}(\Sigma) \\ 1 & \text{pro } ab \in \mathcal{L}(\Sigma) \end{cases}$$

Věta 125 *Nechť Σ je markovský posun s ireducibilní přechodovou maticí M a nechť ϱ je spektrální poloměr M. Pak $\mathcal{H}(\Sigma) = \log \varrho$.*

Důkaz: Slovo $u \in A^n$ náleží do $\mathcal{L}(\Sigma)$ právě když $M(u_0, u_1) \cdots M(u_{n-2}, u_{n-1}) = 1$. Složitost $P(n)$ je tedy součet všech prvků matice M^{n-1}. Z Perron-Frobeniovy věty plyne, že existují konstanty $0 < c_0 < c_1$ takové že $c_0\varrho^n \leq P(n) \leq c_1\varrho^n$. Odtud $\lim_{n\to\infty} \frac{1}{n} \log P(n) = \log \varrho$. □

Příklad 15 *Entropie posunu zlatého řezu je maximum entropií procesů, jejichž nosič je posun zlatého řezu.*

Důkaz: Markovský proces jehož nosič je posun zlatého řezu má přechodovou maticí

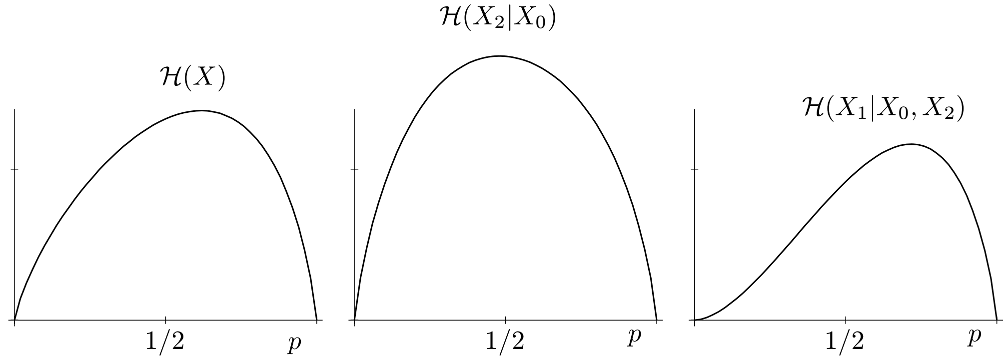
$$R = \begin{bmatrix} p & 1-p \\ 1 & 0 \end{bmatrix}, \quad \pi = \left[\frac{1}{2-p}, \frac{1-p}{2-p}\right], \quad R^2 = \begin{bmatrix} 1-p+p^2 & p-p^2 \\ p & 1-p \end{bmatrix}$$

Entropie je $\mathcal{H}(X) = \mathcal{H}(p, 1-p)/(2-p)$,

$$\mathcal{H}(X_2|X_0) \;=\; \mathcal{H}(1-p+p^2, p-p^2)/(2-p) + \mathcal{H}(p, 1-p)(1-p)/(2-p).$$

Maximální entropii $\mathcal{H}(X) = \log \frac{\sqrt{5}+1}{2} = 0.69$ má proces pro $p = \frac{\sqrt{5}-1}{2} = 0.618$. Pro entropii posunu zlatého řezu dostáváme

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \varrho = \frac{\sqrt{5}+1}{2} = 1.618, \quad H(\Sigma) = 0.69 \;\square$$



Obrázek 22: Entropie markovského procesu

Věta 126 (Parryho věta) *Nechť Σ je markovský posun s ireducibilní přechodovou maticí M. Nechť ϱ je spektrální poloměr M, a u, v odpovídající pravý a levý vlastní vektor, jejichž skalární součin je 1, tj. $Mu^T = \varrho u^T$, $vM = \varrho v$, $vu^T = 1$. Položme*

$$R(a,b) = \frac{M(a,b)u(b)}{\varrho u(a)}, \quad \pi(a) = v(a)u(a)$$

Pak R je stochastická matice π je její stacionární rozdělení. Pro markovský proces X s maticí R platí $\mathcal{H}(X) = \mathcal{H}(\Sigma)$.

Důkaz: Platí

$$\sum_{b\in B} R(a,b) = 1, \quad \sum_{a\in A} \pi(a)R(a,b) = \sum_{a\in A} \frac{v(a)M(a,b)u(b)}{\varrho} = \pi(b)$$

takže R je stochastická matice π je její stacionární rozložení. Pro $a \in A^n$ je

$$\begin{aligned}
\mathbb{P}[X_{[0,n)} = a_{[0,n)}] &= v(a_0)u(a_0)\frac{M(a_0,a_1)u(a_1)}{\varrho u(a_0)} \cdots \frac{M(a_{n-2},a_{n-1})u(a_{n-1})}{\varrho u(a_{n-2})} \\
&= v(a_0)M(a_0,a_1)\cdots M(a_{n-2},a_{n-1})u(a_{n-1})/\varrho^{n-1} \\
&= \begin{cases} v(a_0)u(a_{n-1})\varrho^{-n+1} & \text{pro } a \in \mathcal{L}^n(\Sigma) \\ 0 & \text{pro } a \notin \mathcal{L}^n(\Sigma) \end{cases}
\end{aligned}$$

Položme $c_0 = \min\{\varrho v(a)u(a) : a \in A\}$, $c_1 = \max\{\varrho v(a)u(a) : a \in A\}$, takže

$$\begin{array}{rcl}
c_0\varrho^{-n} \leq & \mathbb{P}[X_{[0,n)} = a] & \leq c_1\varrho^{-n} \\
\sum_{a\in\mathcal{L}^n(\Sigma)} \mathcal{P}_X(a)(n\log\varrho - c_1) \leq & \mathcal{H}(X_{[0,n)}) & \leq \sum_{a\in\mathcal{L}^n(\Sigma)} \mathcal{P}_X(a)(n\log\varrho - c_0) \\
n\log\varrho - c_1 \leq & \mathcal{H}(X_{[0,n)}) & n\log\varrho - c_0
\end{array}$$

a $\mathcal{H}(X) = \log\varrho$. $\square$

Pro informační obsah platí

$$\mathfrak{I}_{X_{[0,n)}} = (n-1)\log\varrho - \log v(X_0) - \log u(X_{n-1})$$

Odtud dostáváme silnější verzi Entropické věty: $\lim_{n\to\infty} \mathfrak{I}_{x_{[0,n)}}/n = \log\varrho = \mathcal{H}(X)$ platí pro každé $x \in \Sigma$.

6.5 Sofické posuny

Definice 41 *Přechodová funkce (konečného automatu) nad abecedou A je zobrazení $\delta : Q \times A \to Q$, kde Q je konečná (stavová) množina. Funkce δ se rozšiřuje na funkci $\delta : Q \times A^* \to Q$*

$$\delta(q, \lambda) = q, \quad \delta(q, ua) = \delta(\delta(q, u), a), \quad u \in A^*, \quad a \in A$$

Konečný automat může rozeznávat jazyky. Automat startuje z vyznačeného počátečního stavu $i \in Q$. Jestliže po přečtení slova $u \in A^*$ skončí v některém odmítajícím stavu, vstupní slovo je odmítnuto, jinak je přijato. Přijímající automat je specifikován čtveřicí (Q, δ, i, R), kde Q je konečná množina stavů, $\delta : Q \times A \to Q$ je přechodová funkce, $i \in Q$ je počáteční stav a $R \subseteq Q$ je množina odmítajících stavů. Jazyk akceptovaný (Q, δ, i, R) je

$$\{u \in A^* : \ \delta(i, u) \notin R\}.$$

Jazyk $L \subseteq A^*$ je regulární, pokud je akceptován nějakým automatem.

Je-li L jazyk posunu, je centrální, tj. obsahuje každé podslovo každého svého slova, a pro každé slovo $u \in L$ existuje $a \in A$ takové že $ua \in L$. Automat který akceptuje centrální regulární jazyk musí splňovat další podmínky. Protože $\lambda \in L$, počáteční stav nemůže být odmítající. Jakmile se automat dostane do množiny odmítajícíh stavů, nemůže jí opustit, protože každé prodloužení odmítnutého slova musí být odmítnuto. To znamená že můžeme redukovat množinu odmítajících stavů do jediného odmítajícího stavu $\{r\}$. Dále můžeme předpokládat, že každý jiný stav je dosažitelný z počátečního stavu, tj.

$$\forall q \in Q \setminus \{r\}, \exists u \in L, \delta(i, u) = q.$$

Pokud by to neplatilo, mohli bychom vynechat všechny stavy, které nejsou dosažitelné, aniž bychom změnili jazyk. Pokud automat splňuje toto omezení, není třeba vyznačovat počáteční stav, protože každý stav kromě r může sloužit jako počáteční. Platí

$$q \in Q \setminus R \ \& \ \delta(q, u) \in Q \setminus R \implies u \in L$$

Protože totiž q je dosažitelný, existuje $v \in L$ takový, že $\delta(i, v) = q$, takže $\delta(i, vu) = \delta(q, u) \in Q \setminus R$, a $vu \in L$. Protože L je centrální, $u \in L$. To zjednodušuje definici přijímajícího automatu.

Definice 42 *Přijímající automat nad abecedou A je trojice (Q, δ, r), kde Q je konečná množina, $r \in Q$ je odmítající stav a $\delta : Q \times A \to Q$ je přechodobvá funkce splňující*

(1) $\forall q \in Q \setminus \{r\}, \exists a \in A, \delta(q, a) \neq r.$

(2) $\forall a \in A, \delta(r, a) = r.$

Jazyk přijímaný (Q, δ, r) je $L = \{u \in A^* : \ \exists q \in Q, \delta(q, u) \neq r\}$. Posun $\Sigma \subseteq A^{\mathbb{N}}$ je sofický, pokud jeho jazyk je přijímaný nějakým přijímajícím automatem.

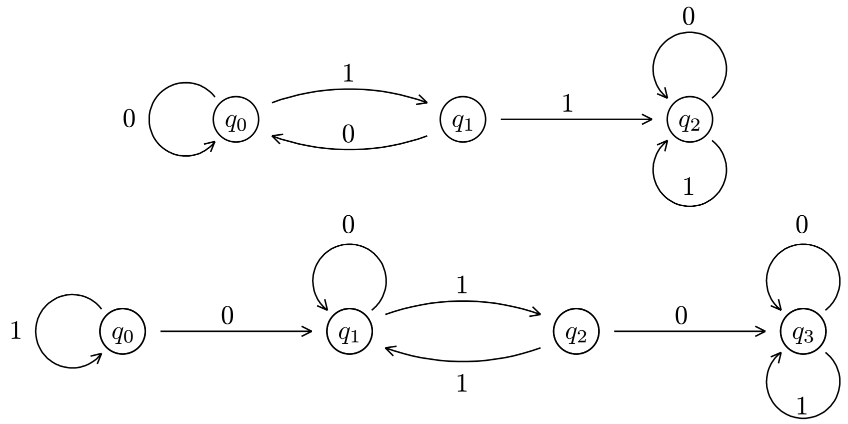
Naopak každý jazyk přijímaný konečným automatem je centrální, takže určuje sofický posun.

Příklad 16 *Posun zlatého řezu a sudý posun jsou sofické.*

Sestrojíme automaty s přechodovými funkcemi

	q_0	q_1	q_2
0	q_0	q_0	q_2
1	q_1	q_2	q_2

	q_0	q_1	q_2	q_3
0	q_1	q_1	q_3	q_3
1	q_0	q_2	q_1	q_3



Obrázek 23: Grafy akceptujících automatů

Pro posun zlatého řezu je $Q = \{q_0, q_1, q_2\}$, $i = q_0$, and $r = q_2$. Automat je ve stavu q_1, pokud poslední přečtené písmeno je 1. Pro sudý posun je $Q = \{q_0, q_1, q_2, q_3\}$, $i = q_0$ a $r = q_3$. Automat zůstává v počátečním stavu q_0 dokud nepřečte nulu. Pak přechází do stavu q_1 a zůstává zde dokud čte další nuly. Po přečtení jedničky vstoupí do q_2. Pokud nyní přečte nulu, vstupní slovo je odmítnuto, jinak se vrací do q_1.

Konečné automaty znázorňujeme orientovanými označenými grafy, jejichž vrcholy jsou stavy z Q a jejichž hrany jsou označeny písmeny A. Existuje hrana z q do q' označená a, právě když $\delta(q, a) = q'$. Z každého vrcholu tedy vede právě $|A|$ hran označených písmeny A (Obrázek 23). V tomto grafu je však odmítající stav zbytečný. Pokud ho vynecháme a vynecháme též všechny hrany které do něj vedou, množina přijímaných slov se nezmění. Slovo odmítneme pokud v grafu nemůžeme pokračovat hranou označenou dalším písmenem slova.

Definice 43

(1) Graf je čtveřice $G = (V, E, s, t)$, kde V je konečná množina vrcholů, E je konečná množina hran, $s, t : E \to V$ jsou zobrazení přiřazující hraně její počáteční a koncový vrchol.

(2) Označený graf nad abecedou A je dvojice (G, l), kde G je graf a $l : E_G \to A$ je označovací zobrazení.

(3) Označený graf (G, l) je rezolventní, pokud

$$\forall e_1, e_2 \in E_G, (s(e_1) = s(e_2) \,\&\, l(e_1) = l(e_2) \implies e_1 = e_2).$$

(3) posun Σ_G grafu G je $\Sigma_G = \{u \in E^{\mathbb{N}} : \forall i \geq 0, t(u_i) = s(u_{i+1})\} \subseteq E^{\mathbb{N}}$.

(4) Posun označeného grafu (G, l) je $\Sigma_{(G,l)} = l(\Sigma_G) \subseteq A^{\mathbb{N}}$.

Zřejmě Σ_G je markovský posun. Zobrazení $l : E^{\mathbb{N}} \to A^{\mathbb{N}}$ definované $l(u)_i = l(u_i)$. je spojité a $\Sigma_{(G,l)} = l(\Sigma_G)$, takže $\Sigma_{(G,l)}$ je posun. Označené grafy pro posun zlatého řezu a sudý posun jsou na obrázku 24.

Tvrzení 127 *Je-li $\Sigma \subseteq A^{\mathbb{N}}$ sofický posun pak existuje rezolventní označený graf (G, l) takový že $\Sigma = \Sigma_{(G,l)}$.*

Důkaz: Nechť (Q, δ, r), je automat který přijímá $\mathcal{L}(\Sigma)$. Sestrojíme graf $G = (V, E, s, t)$, kde

$$V = Q \setminus \{r\}, \quad E = \{(q, a) \in V \times A : \; a \in A, \delta(q, a) \neq r\},$$

$s(q, a) = q$, $t(q, a) = \delta(q, a)$. Označení $l : E \to A$ je $l(q, a) = a$. $\square$

Obrázek 24: Sofické posuny

Tvrzení 128 *Každý posun označeného grafu je sofický.*

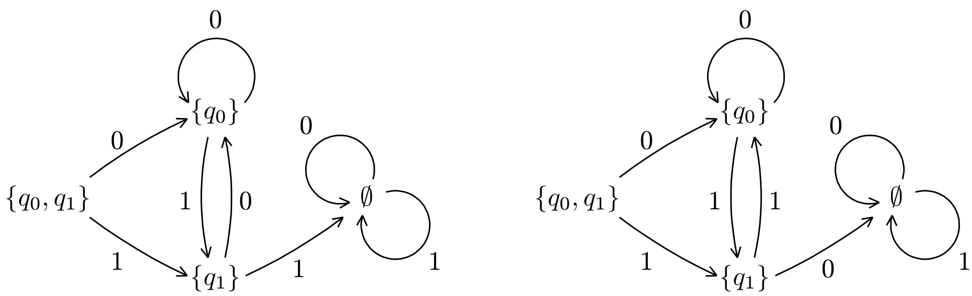
Důkaz: Nechť (G, l) je označený graf, nechť Q je množina podmnožin V_G. Definujem přechodovou funkci $\delta : Q \times A \to Q$

$$\delta(M, a) = \{q \in Q : \exists e \in E_G, s(e) \in M, t(e) = q, l(e) = a\}.$$

Počáteční stav je V a odmítající stav je $\emptyset$. Ukážeme, že $(Q, \delta, V, \emptyset)$ přijímá $\mathcal{L}(\Sigma_{(G,l)})$. Pokud

$$q_0 \xrightarrow{u_0} q_1 \xrightarrow{u_1} q_2 \dots q_{n-1} \xrightarrow{u_{n-1}} q_n$$

je cesta v (G, l), pak $q_n \in \delta(V, u)$, takže $\delta(V, u) \neq \emptyset$ a u je přijímáno. Naopak pokud $\delta(V, u) \neq \emptyset$, vybereme nějaký $q_n \in \delta(V, u)$. Existuje $q_{n-1} \in \delta(V, u_{[0,n-2]})$ takový že $q_{n-1} \xrightarrow{u_{n-1}} q_n$ je označená hrana v G. Takto pokračujeme (pozpátku) a získáme cestu v G s označením u. $\square$



Obrázek 25: Přijímající automaty označených grafů

Každý sofický posun je faktor posunu konečného typu, protože $l : \Sigma_G \to l(\Sigma_G)$ je faktorizace (surjektivní okénkový kód). Naopak platí

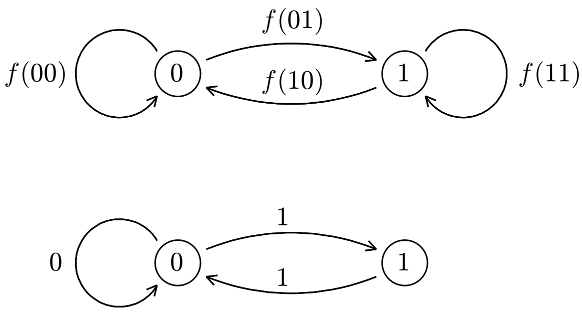
Věta 129 (Weiss) *Posun je sofický právě když je faktorem posunu konečného typu.*

Důkaz: Pokud Σ je sofický, pak $\Sigma = \Sigma_{(G,l)}$ pro nějaký označený graf (G, l) a $l : \Sigma_G \to \Sigma_{(G,l)}$ je faktorizace. Naopak nechť $F : (\Sigma, \sigma) \to (\Theta, \sigma)$ je surjektivní, $\Sigma \subseteq A^{\mathbb{N}}$ SFT a $\Theta \subseteq B^{\mathbb{N}}$. Můžeme předpokládat, že Σ je markovský posun. Podle Hedlundovy věty 123 existuje lokální pravidlo $f : \mathcal{L}^{r+1}(\Sigma) \to B$ takové že $F(x)_i = f(x_{[i,i+r]})$. Můžeme předpokládat $r \geq 1$. Definujme označený graf (G, l), kde $V_G = \mathcal{L}^r(\Sigma)$, $E_G = \mathcal{L}^{r+1}(\Sigma)$,

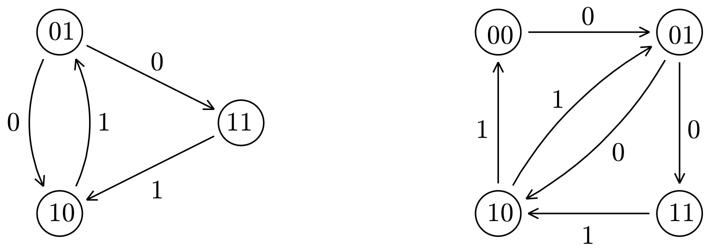
$$\begin{aligned} s(u_0 u_1, \dots u_{r-1} u_r) &= u_0 \dots u_{r-1} \\ t(u_0 u_1, \dots u_{r-1} u_r) &= u_1 \dots u_r \end{aligned}$$

Pak $\Sigma_{(G,l)} = \Sigma$. $\square$

Entropii posunu lze určit z matice přiřazené jejímu grafu.



Obrázek 26: Faktory SFT



Obrázek 27: Grafy $\Sigma_{\{00,111\}}$ a $\Sigma_{\{000,111\}}$

Definice 44 *Matice $M = (M_{ab})_{a,b \in V}$ grafu $G = (V, E, s, t)$ je*

$$M_{ab} = |\{e \in E : \; s(e) = a, t(e) = b\}|.$$

Tvrzení 130 *Nechť $G = (V, E, s, t)$ je graf a λ je spektrální poloměr jeho matice. Pak $\mathcal{H}(\Sigma_G) = \log \lambda$.*

Důkaz: Pro každé $a, b \in V$ a $n > 0$ platí

$$(M^n)_{ab} = \sum_{c_1,\ldots,c_{n-1}} M_{a,c_1} \cdots M_{c_{n-1}b} = |\{u \in \mathcal{L}^n(\Sigma_G) : \; s(u_0) = a, \; t(u_{n-1}) = b\}|$$

Pro složitost dostáváme $P_{\Sigma_G}(n) = \sum_{a,b \in V} (M^n)_{ab}$, takže $\mathcal{H}(\Sigma_G) = \log \lambda$. $\square$

Věta 131 *Nechť (G, l) je rezolventní označený graf a λ je spektrální poloměr matice G. Pak $\mathcal{H}(\Sigma_{(G,l)}) = \log \lambda$.*

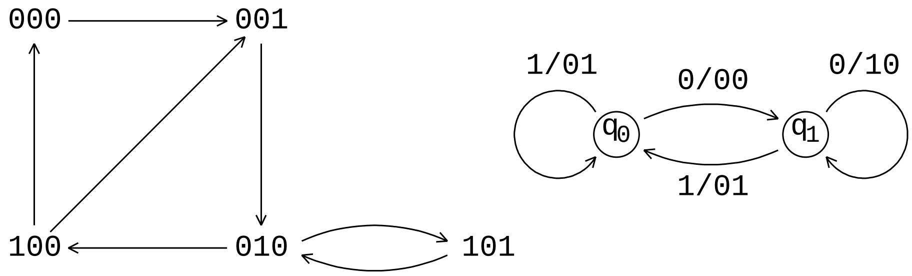
Důkaz: Nechť $k = |V_G|$. Pro každé $u \in \mathcal{L}(\Sigma_{(G,l)})$ existuje nejvýše k slov $v \in \mathcal{L}(\Sigma_G)$ pro která $l(v) = u$. Odtud

$$P_{\Sigma_G}(n) \le P_{\Sigma_{(G,l)}}(n) \le k \cdot P_{\Sigma_G}(n)$$

a $\mathcal{H}(\Sigma_{(G,l)}) = \mathcal{H}(\Sigma_G)$. $\square$

6.6 Automatické kódy

Pro posun $\Sigma_{(1,3)} = \Sigma_{\{0000,11\}}$ je graf posunu $\Sigma^{(3)}_{(1,3)}$ na obrázku 28 vlevo. Z matice tohoto grafu lze určit spektrální poloměr $\varrho = 1.465$, takže jeho entropie je $\mathcal{H}(\Sigma_{(1,3)}) = 0.551$. Protože $2 \cdot \mathcal{H}(\Sigma_{1,3)}) > 1$, lze binární slova kódovat v tomto posunu dvojnásobnou délkou. Takový kód lze realizovat konečným automatem na obrázku 28 vpravo. Počáteční stav automatu je



Obrázek 28: Automatický kód

q_0. Při vstupu $a \in \{0,1\}$ automat v prvním kroku sleduje hranu s označením a/b a na výstup zapíše slovo $b \in \{0,1\}^2$. Každé $x \in A^{\mathbb{N}}$ je pak kódováno nějakým $y \in \Sigma_{(1,3)}$ a toto zobrazení je prosté.

Definice 45 *Konečný automat z abecedy A do abecedy B je trojice $\mathcal{A} = (Q, \delta, \eta)$, kde $\delta : Q \times A \to Q$ je přechodová funkce a $\eta : Q \times A \to B^*$ je výstupní funkce, taková že pro každé $q \in Q$, $u \in A^+$ platí*

$$\delta(q,u) = q \implies \eta(q,u) \in B^+.$$

Přechodová a výstupní funkce se rozšiřuje na zobrazení $\delta : Q \times A^* \to Q$, $\eta : Q \times A^* \to B^*$, $\eta : Q \times A^{\mathbb{N}} \to B^{\mathbb{N}}$ rekurentním předpisem

$$\delta(q,ua) = \delta(\delta(q,u),a), \quad \eta(q,au) = \eta(q,a)\eta(\delta(q,a),u)$$

Pro $q \in Q$ označme $\eta_q : A^* \to B^*$ a $\eta_q : A^{\mathbb{N}} \to B^{\mathbb{N}}$ zobrazení daná předpisem $\eta_q(u) = \eta(q,u)$.

Tvrzení 132 *Je-li (Q,δ,η) konečný automat, je každé zobrazení η_q spojité.*

Důkaz: Je-li $u \in A^*$ a $|u| \ge |Q|$, je $|\eta(q,u)| > 0$. z toho plyne

$$\begin{aligned} d(x,y) < 2^{-n|Q|} &\implies x_{[0,n|Q|]} = y_{[0,n|Q|]} \implies \eta_q(x)_{[0,n]} = \eta_q(y)_{[0,n]} \\ &\implies d(\eta_q(x), \eta_q(y)) < 2^{-n} \;\square \end{aligned}$$

Definice 46 *Zobrazení $F : A^{\mathbb{N}} \to B^{\mathbb{N}}$ je automatický kód, pokud existuje automat (Q,δ,η) a $q \in Q$ takový že $F = \eta_q$. Zobrazení $F : A^{\mathbb{N}} \to B^{\mathbb{N}}$ je uzavírací, pokud existuje $m > 0$ takové, že pro všechna $x, y \in A^{\mathbb{N}}$ platí*

$$F(x)_{[0,n+m)} = F(y)_{[0,n+m)} \implies x_{[0,n)} = y_{[0,n)}$$

Speciálním případem automatických kódů jsou blokové kódy, které jsou definovány zobrazením $f : A^n \to B^*$ předpisem $F(x) = f(x_{[0,n)})f(x_{[n,2n)})\cdots$. Okénkové kódy (celulární automaty) jsou definovány zobrazením $f : A^{d+1} \to B$ předpisem $F(x)_i = f(x_{[i,i+d]})$. Každé uzavírací zobrazení je prosté. Naopak platí

Tvrzení 133 *Nechť $F : A^{\mathbb{N}} \to B^{\mathbb{N}}$ je prostý automatický kód. Pak F je uzavírací a inverzní kód $F^{-1} : B^{\mathbb{N}} \to A^{\mathbb{N}}$ je také automatický.*

Důkaz: Nechť (Q,δ,η) je automat, $q_0 \in Q$ a předpokládejme sporem, že $F = \eta_{q_0}$ je prosté a není uzavírací. Nechť $Q_0 = \{q \in Q : \exists u \in A^*, \delta(q_0,u) = q\}$ je množina stavů dosažitelných z q_0. Pro $m > 0$ položme

$$Y_m = \{(q,x,y) \in Q_0 \times A^{\mathbb{N}} \times A^{\mathbb{N}} : \; x_0 \ne y_0, \;\; \eta(q,x)_{[0,m]} = \eta(q,y)_{[0,m]}\}$$

Ukážeme, že množiny Y_m jsou neprázdné. Podle předpokladu pro každé $m > 0$ existují x, y a $n \geq 0$ takové že $\eta(q_0, x)_{[0,n+m]} = \eta(q_0, y)_{[0,n+m]}$ a $x_{[0,n]} \neq y_{[0,n]}$. Je-li $i < n$ první index pro který $x_i \neq y_i$, je $\sigma^i(x)_0 \neq \sigma^i(y)_i$ a pro $q = \delta(q_0, x_{[0,i)}) \in Q_0$ platí $\eta(q, \sigma^i(x))_{[0,m]} = \eta(q, \sigma^i(y))_{[0,m]}$, takže $(q, \sigma^i(x), \sigma^i(y)) \in Y_m$. Množiny Y_m jsou tedy uzavřené neprázdné množiny v kompaktním prostoru $Q_0 \times A^{\mathbb{N}} \times A^{\mathbb{N}}$. Protože $Y_{m+1} \subseteq Y_m$, je průnik všech Y_m také neprázdný. Jestliže (q, x, y) náleží do tohoto průniku, existuje $i \geq 0$ a $u \in A^i$, pro které $\delta(q_0, u) = q$. Pro nekonečná slova ux, uy dostáváme $(ux)_i \neq (uy)_i$ a $\eta(q_0, ux) = \eta(q_0, uy)$ a to je spor, takže F je uzavírací. Odtud plyne, že pro každé $q \in Q_0$ je η_q uzavírací. Ukážeme že inverzní zobrazení F^{-1} je také automatické. Pro $q \in Q_0$ označme m_q uzavírací konstantu zobrazení η_q. Existuje tedy zobrazení $f_q : B^{m_q} \to A$ takové že platí

$$\eta_q(x)_{[0,m_q)} = \eta_q(y)_{[0,m_q)} \implies x_0 = y_0$$

Pro konstrukci inverzního automatu položme $Q' = Q_0 \times B^*$, a definujme $\delta' : Q' \times B \to Q'$, $\eta' : Q' \times B \to A^*$ náseldujícím způsobem. Pro dané $(q, u) \in Q'$ nechť k je nejmenší číslo pro které platí $|u| - k < m_{\delta(q, u_{[0,k)})}$. Pak

$$\begin{aligned} \delta'((q, u), b) &= (\delta(q, u_{[0,k)}), \sigma^k(u)b) \\ \eta'((q, u), b)_i &= f_{q'}(u_{[i,i+m_{q'})}), \text{ kde } q' = \delta(q, u_{[0,i)}), \ \ i < k \end{aligned}$$

Počáteční stav je $q'_0 = (q_0, \lambda)$ a $Q'_0 \subseteq Q'$ je konečná množina stavů dosažitelných z q'_0. Pak (Q'_0, δ', η') je konečný automat a $F^{-1} = \eta'_{q'_0}$. □

Tvrzení 134 *Je-li $F : A^{\mathbb{N}} \to B^{\mathbb{N}}$ prostý automatický kód a $\Sigma \subseteq B^{\mathbb{N}}$ posun takový že $F(A^{\mathbb{N}}) \subseteq \Sigma$, pak $\mathcal{H}(\Sigma) \geq \log |A|$.*

Důkaz: Podle Tvrzení 133 je F uzavírací. Je-li m uzavírací konstanta, je $P_\Sigma(n+m) \geq |A|^n$, takže $\mathcal{H}(\Sigma) \geq \log |A|$. □

Věta 135 *Nechť $\Sigma \subseteq B^{\mathbb{N}}$ je sofický posun. Existuje automatický uzavírací kód $F : A^{\mathbb{N}} \to \Sigma$ právě když $\log |A| \leq \mathcal{H}(\Sigma)$.*

Důkaz viz Lind a Marcus [20].

7 Přenos informace

7.1 Fanova nerovnost

Říkáme, že náhodná veličina $X : \Omega \to A$ je určena náhodnou veličinou $Y : \Omega \to B$, pokud existuje funkce $f : B \to A$ taková že $\mathbb{P}[X = f(Y)] = 1$.

Tvrzení 136 *Náhodná veličina $X : \Omega \to A$ je určena náhodnou veličinou $Y : \Omega \to B$ právě když $\mathcal{H}(X|Y) = 0$.*

Důkaz: Podle definice je $\mathcal{H}(X|Y) = \sum_{b \in B} \mathcal{H}(X|Y = b) \cdot \mathbb{P}[Y = b]$. Je-li $\mathcal{H}(X|Y) = 0$, musí být každé $\mathcal{H}(X|Y = b)$ nulové, takže příslušný sloupec matice podmíněných pravděpodobností obsahuje jedinou jednotku. To znamená že existuje $a = f(b)$ pro které $\mathbb{P}[X = f(b)|Y = b] = 1$ a tedy také $\mathbb{P}[X = f(Y)] = 1$. Naopak je-li $\mathbb{P}[X = f(Y)] = 1$, je $\mathcal{H}(X|Y = b) = 0$ pro každé $b \in B$ a tedy také $\mathcal{H}(X|Y) = 0$. $\square$

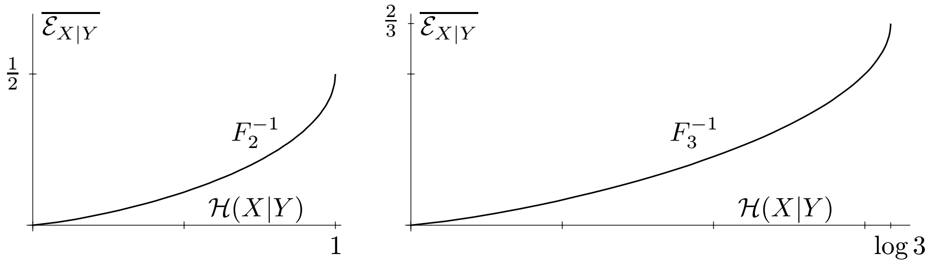
Je-li $\mathcal{H}(X|Y) > 0$, můžeme X odhadovat z Y pomocí funkce odhadu $f : B \to A$. Definujme pravděpodobnost chyby odhadu $a \in A$ a průměrnou pravděpodobnost chyby odhadu vzorci

$$\begin{aligned}
\mathcal{E}_{X|Y}(f, a) &= \mathbb{P}[f(Y) \neq a | X = a] = \sum_{b \in B \setminus f^{-1}(a)} P_{XY}(a, b)/P_X(a) \\
\overline{\mathcal{E}_{X|Y}}(f) &= \mathbb{P}[f(Y) \neq X] = \sum_{a \in A} \mathcal{E}_{X|Y}(f, a) \cdot P_X(a) = 1 - \sum_{b \in B} P_{XY}(f(b), b) \\
\overline{\mathcal{E}_{X|Y}} &= \min\{\overline{\mathcal{E}_{X|Y}}(f) : \ f : B \to A\}
\end{aligned}$$

Pro $x \in [0, 1]$ položme $h(x) = \mathcal{H}(x, 1 - x) = -x \cdot \log(x) - (1 - x) \cdot \log(1 - x)$ a

$$F_n(x) = h(x) + x \cdot \log(n - 1).$$

Funkce F_n je rostoucí v intervalu $[0, 1 - \frac{1}{n}]$, má minimum $F_n(0) = 0$ a maximum $F_n(1 - \frac{1}{n}) = \log n$. Inverzní funkce $F_n^{-1} : [0, \log n] \to [0, 1 - \frac{1}{n}]$ je na Obrázku 29



Obrázek 29: Fanova nerovnost

Tvrzení 137 *Nechť $X : \Omega \to A$, $Y : \Omega \to B$ jsou náhodné proměnné. Pak*

$$\overline{\mathcal{E}_{X|Y}} \geq F_{|A|}^{-1}(\mathcal{H}(X|Y)).$$

Důkaz: Nechť $f : B \to A$ je funkce odhadu a nechť $E = [f(Y) \neq X]$ je náhodná proměnná "odhad je chybný", tj.

$$E(\omega) = \begin{cases} 0 & \text{pokud} \quad f(Y(\omega)) = X(\omega) \\ 1 & \text{pokud} \quad f(Y(\omega)) \neq X(\omega). \end{cases}$$

Pak $\mathcal{H}(E) = h(\overline{\mathcal{E}_{X|Y}}(f))$, $\mathcal{H}(E|X,Y) = 0$, $\mathcal{H}(X|Y, E = 0) = 0$, $\mathcal{H}(X|Y, E = 1) \le \log(|A| - 1)$, takže

$$\begin{aligned}
\mathcal{H}(X|Y) &= \mathcal{H}(X|Y) + \mathcal{H}(E|X,Y) = \mathcal{H}(E,X|Y) = \mathcal{H}(E|Y) + \mathcal{H}(X|Y,E) \\
&\le \mathcal{H}(E) + \mathcal{H}(X|Y,E=0)\cdot \mathbb{P}[E=0] + \mathcal{H}(X|Y,E=1)\cdot \mathbb{P}[E=1] \\
&\le h(\overline{\mathcal{E}_{X|Y}}(f)) + \overline{\mathcal{E}_{X|Y}}(f)(|A|-1) = F_n(\overline{\mathcal{E}_{X|Y}}(f))
\end{aligned}$$

Protože tato nerovnost platí pro každé $f : B \to A$, platí také pro $\overline{\mathcal{E}}_{X|Y}$. □

Odhad Fanovy nerovnosti nelze zlepšit. Předpokládejme, že X, Y jsou nezávislé, tj. $\mathcal{H}(X|Y) = \mathcal{H}(X)$. V tomto případě hodnota X musí být odhadnuta bez jakékoliv předběžné informace, a nejlepší odhad je hodnota X s největší pravděpodobností. Je-li $|A| = n \ge 2$ a $P_X = (1-r, \frac{r}{n-1}, \ldots, \frac{r}{n-1})$, kde $1 - r > r/(n-1)$, je nejlepší odhad $f(X) = 0$ a $\overline{\mathcal{E}_{X|Y}} = r$. Ve Fanově nerovnosti pak dostáváme rovnost $f_n(r) = h(r) + r\log(n-1) = \mathcal{H}(P_X)$.

7.2 Kapacita informačního kanálu

Informační kanál je model zařízení, které přenáší zprávy. Při přenosu dochází k chybám, ale známe pravděpodobnostní rozložení těchto chyb. Pro každé písmeno vstupní abecedy $a \in A$ známe pravděpodobnostní rozložení na výstupní abecedě B.

Definice 47 *Informační kanál $R : A \to B$ je stochastická matice $R = (R(a,b))_{a\in A, b\in B}$, tj. matice nezáporných čísel která pro každé $a \in A$ splňuje $\sum_{b\in B} R(a,b) = 1$.*

Každý řádek matice R je tedy pravděpodobnostní rozdělení na abecedě B. Přichází-li na vstup náhodná proměnná X s rozdělením $P_X = P$, je rozložení vstupní náhodné veličiny $P_Y = P \cdot R$, tj. součin řádkového vektoru P s maticí R, tedy $P_Y(b) = \sum_{a\in A} P(a)\cdot R(a,b)$. Rozdělení dvojice (X,Y) je matice kterou značíme $P \times R$, tj. $(P\times R)(a,b) = P(a)\cdot R(a,b)$. Množství přenesené informace, tj. vzájemná informace $\mathcal{I}(X : Y)$ závisí pouze na rozdělení X. Kapacita kanálu je maximální možné množství vzájemné informace.

Definice 48 *Kapacita informačního kanálu R je*

$$\begin{aligned}
C(R) &= \max\{\mathcal{I}(X:Y) :\ \mathbb{P}[Y=b|X=a] = R(a,b)\} \\
&= \max\{\mathcal{H}(P) + \mathcal{H}(P\cdot R) - \mathcal{H}(P\times R) :\ P \in \Delta(A)\}
\end{aligned}$$

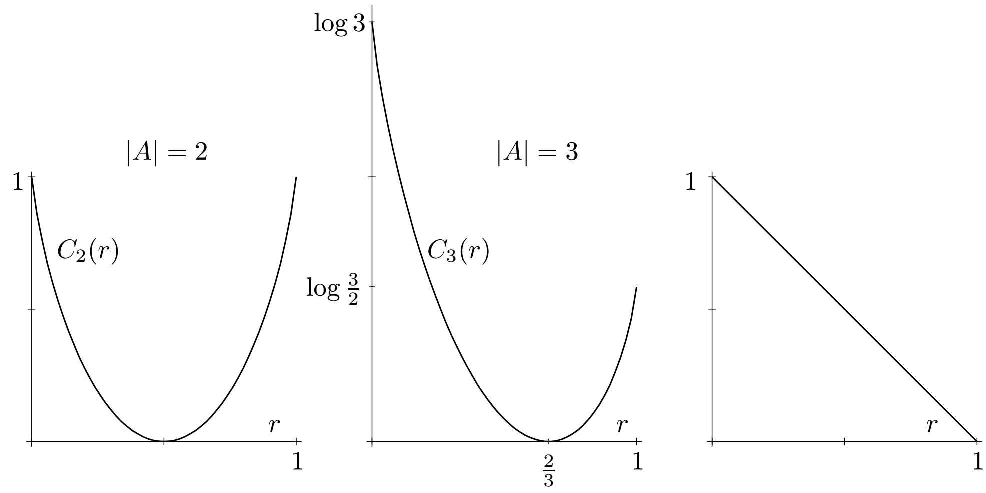
Příklad 17 *Kapacita binárního symetrického kanálu (BSC)*

$$R = \begin{bmatrix} 1-r & r \\ r & 1-r \end{bmatrix} \ \text{je}\ C(R) = 1 - h(r).$$

Důkaz: Při výpočtu kapacity využijeme toho, že oba řádky matice R mají stejnou entropii. Jsou-li X, Y náhodné proměnné s podmíněnou pravděpodobností R, je

$$\begin{aligned}
\mathcal{I}(X:Y) &= \mathcal{H}(Y) - \mathcal{H}(Y|X) = \mathcal{H}(Y) - \sum_{a\in A} P_A(a)\mathcal{H}(Y|X=a) \\
&= \mathcal{H}(Y) - h(r) \le 1 - h(r)
\end{aligned}$$

Tento horní odhad na $\mathcal{I}(X : Y)$ je nabýván při rovnoměrném rozdělení X, při kterém je rozdělení Y také rovnoměrné (Obrázek 30 vlevo). □



Obrázek 30: Kapacita informačního kanálu

Příklad 18 *Kapacita q-árního symetrického kanálu (Obrázek 30 uprostřed)*

$$R = \begin{bmatrix} 1-r & r/(q-1) & \cdots & r/(q-1) \\ & \vdots & & \\ r/(q-1) & \cdots & r/(q-1) & 1-r \end{bmatrix}$$

je

$$C_q(r) = \log q - \mathcal{H}(1-r, \tfrac{r}{q-1}, \ldots, \tfrac{r}{q-1}) = \log q + (1-r)\log(1-r) + r\log\tfrac{r}{q-1}.$$

Důkaz: Při výpočtu kapacity opět využijeme toho, že každý řádek matice R má stejnou entropii. Jsou-li X, Y náhodné proměnné s podmíněnou pravděpodobností R, je

$$\begin{aligned} \mathcal{I}(X:Y) &= \mathcal{H}(Y) - \mathcal{H}(Y|X) = \mathcal{H}(Y) - \sum_{a\in A} P_X(a)\cdot\mathcal{H}(Y|X=a) \\ &= \mathcal{H}(Y) - \mathcal{H}(1-r, \tfrac{r}{q-1}, \ldots, \tfrac{r}{q-1}) \le \log q - \mathcal{H}(1-r, \tfrac{r}{q-1}, \ldots, \tfrac{r}{q-1}) \end{aligned}$$

Tento horní odhad na $\mathcal{I}(X:Y)$ je nabýván při rovnoměrném rozdělení X, při kterém je rozdělení Y také rovnoměrné. □

Pro $r = 1/2$ má BSC nulovou kapacitu. Pro kapacitu q-árního symetrického kanálu platí

$$C(0) = \log q,\ \ C(\tfrac{q-1}{q}) = 0,\ \ C(1) = \log\tfrac{q}{q-1}.$$

Při přenosu zapomětlivým kanálem se některá vstupní písmena ztrácí, ostatní jsou přenesena bezchybně:

Příklad 19 *Kapacita binární zapomětlivého kanálu (Obrázek 30 vpravo)*

$$R = \begin{bmatrix} 1-r & 0 & r \\ 0 & 1-r & r \end{bmatrix} \ \ \textit{je}\ \ C(R) = 1-r$$

Důkaz: Vzájemná entropie je opět $\mathcal{I}(X:Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) = \mathcal{H}(Y) - h(r)$. Pro rozložení $P_X = (p, 1-p)$ je $P_Y = (p(1-r), (1-p)(1-r), r)$. Označme $E = [X \neq Y]$ náhodnou proměnnou "došlo k chybě". Pak

$$\begin{aligned} \mathcal{H}(Y) &= \mathcal{H}(E) + \mathcal{H}(Y|E) = h(r) + (1-r)\cdot\mathcal{H}(Y|E=0) + r\cdot\mathcal{H}(Y|E=1) \\ &= h(r) + (1-r)h(p) + r\cdot 0 \le h(r) + (1-r) \end{aligned}$$

a toto maximum je nabýváno pro rovnoměrné rozdělení X, kdy $P_Y = (\frac{1}{2}(1-r), \frac{1}{2}(1-r), r)$ a $\mathcal{H}(P_Y) = 1-r+h(r)$. Kapacita zapomětlivého kanálu je tedy $C(R) = h(r)+(1-r)-h(r) = 1-r$. $\square$

Příklad 20 *Kapacita binárního asymetrického kanálu*

$$R = \begin{bmatrix} 1-r & r \\ s & 1-s \end{bmatrix} \quad je \ C(R) = h(\tfrac{1}{a+1}) - p_0 \cdot h(r) - p_1 \cdot h(s), \quad kde$$

$$\log a = \frac{h(r)-h(s)}{1-r-s}, \ \ p_0 = \frac{1-s-as}{(a+1)(1-r-s)}, \ \ p_1 = \frac{a-r-ar}{(a+1)(1-r-s)}$$

Důkaz: Pro funkci $g(x) = -x \log x$ platí $g(xy) = xg(y) + yg(x)$ a $h(x) = g(x) + g(1-x)$. Pro vstupní rozdělení $P = (p_0, p_1) = (p, 1-p)$ je

$$P \times R = \begin{bmatrix} p(1-r) & pr \\ (1-p)s & (1-p)(1-s) \end{bmatrix}, \quad PR = [p(1-r) + (1-p)s, \ \ pr + (1-p)(1-s)]$$

$$\begin{aligned} \mathcal{H}(X,Y) &= p \cdot g(1-r) + p \cdot g(r) + (1-p) \cdot g(s) + (1-p) \cdot g(1-s) \\ &+ g(p) \cdot (1-r) + g(p) \cdot r + g(1-p) \cdot s + g(1-p) \cdot (1-s) \\ &= p \cdot h(r) + (1-p) \cdot h(s) + h(p) \end{aligned}$$

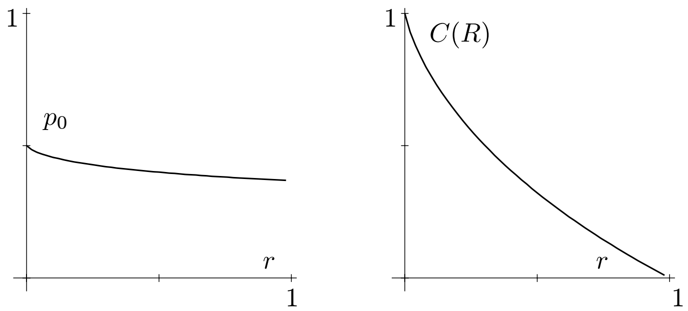
Protože $\mathcal{H}(X) = h(p)$, $\mathcal{H}(Y) = h(p(1-r) + (1-p)s)$, $h'(x) = \log \frac{1-x}{x}$, je

$$\begin{aligned} \mathcal{I}(X:Y) &= h(p(1-r) + (1-p)s) - p \cdot h(r) - (1-p) \cdot h(s) \\ \frac{\partial \mathcal{I}}{\partial p} &= (1-r-s) \cdot \log \frac{pr + (1-p)(1-s)}{p(1-r) + (1-p)s} - h(r) + h(s) \end{aligned}$$

$\partial \mathcal{I}(X:Y)/\partial p = 0$ nastává při $\frac{pr+(1-p)(1-s)}{p(1-r)+(1-p)s} = 2^{\frac{h(r)-h(s)}{1-r-s}} = a$. Odtud již lze vypočítat p jako funkci a a dosazením do vzorce pro $\mathcal{H}(X:Y)$ získáme $C(R)$. $\square$
Speciálně pro $s = 0$ platí

$$\log a = \frac{h(r)}{1-r}, \ \ p_0 = \frac{1}{(a+1)(1-r)}, \ \ C(R) = h(\tfrac{1}{a+1}) - p_0 \cdot h(r)$$

Na obrázku 31 je graf závislosti hodnoty p_0 (vlevo) a $C(R)$ (vpravo) na r při $s = 0$.



Obrázek 31: Kapacita asymetrického kanálu

7.3 Chyba přenosu

Definice 49 *Odhad pro kanál $R : A \to B$ je funkce $f : B \to A$. Je-li $X : \Omega \to A$ vstupní náhodná veličina s rozdělením P_X, je pravděpodobnost chyby*

$$
\begin{aligned}
\mathcal{E}(R,f,a) &= \sum_{b\in B\setminus f^{-1}(a)} R(a,b) = 1 - \sum_{b\in f^{-1}(a)} R(a,b)\\
\overline{\mathcal{E}_X}(f) &= \sum_{a\in A} \mathcal{E}(f,a)\cdot P_A(a) = 1 - \sum_{b\in B} P_A(f(b))\cdot R(f(b),b)\\
\mathcal{E}_X(f) &= \max\{\mathcal{E}(R,f,a) :\ P_X(a) > 0\}\\
\overline{\mathcal{E}_X}(R) &= \min\{\overline{\mathcal{E}_X}(f) :\ f : B \to A\}\\
\mathcal{E}_X(R) &= \min\{\mathcal{E}_X(f) :\ f : B \to A\}
\end{aligned}
$$

Průměrná chyba je vždy nejvýše rovna maximální chybě, tj. $\overline{\mathcal{E}_X}(R) \le \mathcal{E}_X(R)$. Pravděpodobnost chyby lze snížit, posíláme-li zprávy po blocích, tj. používáme místo daného kanálu jeho mocninu.

Definice 50 *Nechť $R : A \to B$ je informační kanál. Jeho n-tá mocnina je informační kanál $R_n : A^n \to B^n$, kde $R_n(u,v) = R(u_0,v_0)\cdots R(u_{n-1},v_{n-1})$ pro $(u,v) \in A^n \times B^n$.*

Kapacita R_n je zřejmě $C(R^n) = n \cdot C(R)$.

Pro binární symetrický kanál R s parametrem $r < 1/2$ je $\mathcal{E}_X(R) = \overline{\mathcal{E}_X}(R) = r$ kdykoliv $P_X(0)$ i $P_X(1)$ jsou kladné. V triviálním případě $P_X(0) = 1$ je sice $\mathcal{E}_X = 0$, pak je ovšem také $\mathcal{H}(X) = 0$. Pravděpodobnost chyby lze snížit, pokud informaci přenášíme po blocích. Je-li kapacita kanálu kladná, lze jím přenášet zprávy s libovolnou přesností. Pro BSC kódujeme každý bit trojicí stejných bitů, tj. $0 \mapsto 000$, $1 \mapsto 111$. Na výstupu pak dekódujeme hlasováním. Je-li v přijaté trojici více jednotek než nul, dekódujeme tuto trojici jako 1. To znamená, že vstupní náhodná proměnná X_3 má rovnoměrné rozdělení na množině $\{000, 111\}$ a pravděpodobnost chyby je $\mathcal{E}_{X_3}(R^3) = r^3 + 3r^2(1-r) \approx 3r^2$. Rychlost přenosu je však $\mathcal{H}(X_3)/3 = 1/3$. Kódujeme-li každý bit pěticí bitů stejné hodnoty, je pravděpodobnost chyby $r^5 + 5r^4(1-r) + 10r^3(1-r)^2 \approx 10r^3$. Při tomto kódování sice při rostoucí délce kódu klesá pravděpodobnost chyby, ale také klesá rychlost přenosu. V prvním případě je 1/3 a v druhém případě 1/5 bitů na znak. Stačí-li nám chyba řádu r^2, uvažujeme čtyřpísmenovou abecedu $A = \{a,b,c,d\}$ kterou kódujeme binárními slovy $\{00000, 00111, 11100, 11011\}$, takže rychlost přenosu je $\log(4)/5 = 2/5$. Dojde-li nyní při přenosu nejvýše k jedné chybě, jsme schopni ji odstranit, takže pravděpodobnost chyby je řádu $5r^2$. Tento postup lze zobecnit

Definice 51 *Hammingova vzdálenost na množině A^n je dána vzorcem*

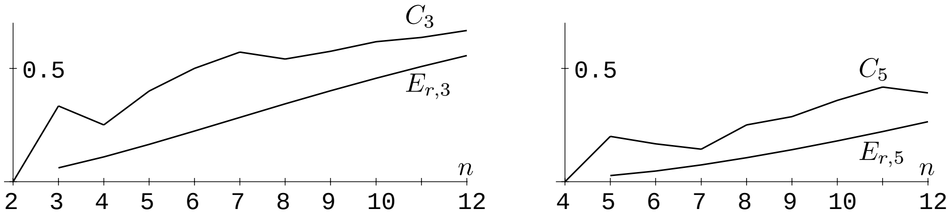
$$d(u,v) = |\{i < n :\ u_i \ne v_i\}|,\ \ B_a(u) = \{b \in A^n :\ d(v,u) \le a\}$$

Množina $K \subseteq A^n$ je a-separovaná, pokud každé dva různé prvky K mají vzdálenost nejméně a. Položme

$$k_a(n) = \max\{|K| :\ K \subseteq A^n \text{ je } a\text{-separovaná}\},\ \ C_a(n) = \log k_a(n)/n$$

Je-li $K \subseteq A^n$ $(2a+1)$-separovaná množina, je $B_a(u) \cap B_a(v) = \emptyset$ pro každé $u \ne v \in A^n$. Množiny $\{000, 111\}$ a $\{00000, 00111, 11100, 11011\}$ jsou 3-separované, takže $k_3(3) = 2$ a $k_3(5) = 4$.

n	3	4	5	6	7	8	9	10	11	12
$k_3(n)$	2	2	4	8	16	20	36	72	128	256
$k_5(n)$	1	1	2	2	2	4	6	12	24	26
$C_3(n)$	0.33	0.25	0.4	0.5	0.57	0.54	0.57	0.62	0.64	0.67
$C_5(n)$	0	0	0.2	0.17	0.14	0.25	0.29	0.36	0.42	0.39



Obrázek 32: Rychlost přenosu a kapacita

Jestliže $K_n \subseteq A^n$ je $(2a+1)$-separovaná množina a X_n náhodná proměnná s rovnoměrným rozložením na K_n, je rychlost přenosu a pravděpodobnost chyby

$$C_{2a+1}(n) = \mathcal{H}(X_n)/n = \log k_n/n, \quad E_{r,2a+1}(n) = \mathcal{E}_{X_n}(R^n) = 1 - \sum_{i=0}^{a} \binom{n}{i} r^i(1-r)^{n-i}$$

Tato závislost je znázorněna na obrázku 32 pro $r = 0.15$. Při rostoucím n roste kapacita ale také pravděpodobnost chyby. Chceme-li přenášet zprávy s malou pravděpodobností chyby, nemůžeme je přenášet rychleji než $C(R)$ bitů na jeden takt. To lze vidět z následujícího heuristického argumentu. Uvažujme BSC s parametrem $0 < r < \frac{1}{2}$. Střední hodnota počtu chyb při přenosu slova délky n je rn. Z Čebyševovy nerovnosti plyne, že pravděpodobnost více než $n(r+\varepsilon)$ chyb konverguje k nule pro $n \to \infty$. Malé pravděpodobnosti chyby lze tedy dosáhnout s kódem $K_m \subseteq A^n$, pro který jsou množiny $\{B_{nr}(u) : \ u \in K_n\}$ navzájem disjunktní. Počet prvků koule s poloměrem nr lze aproximovat

$$B_{nr}(u) = 1 + n + \cdots + \binom{n}{nr} \approx \binom{n}{nr} \approx 2^{n \cdot h(r)}$$

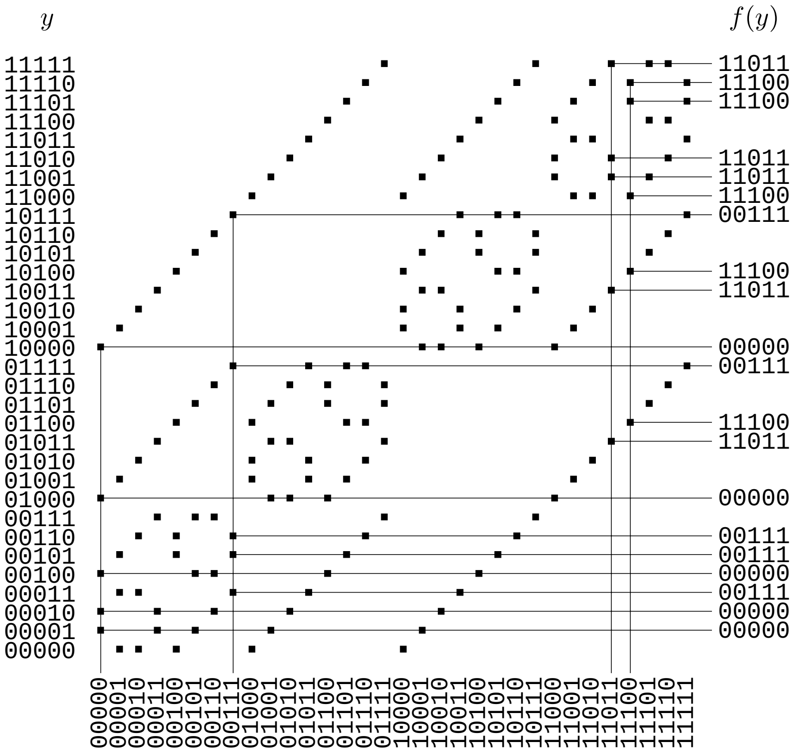
Odtud $|K_n| \approx 2^n/2^{n\cdot h(r)} = 2^{n\cdot C(R)}$, takže $\log |K_n|/n \approx C(R)$. Tento heuristický argument dává horní mez na rychlost přenosu. Shannonova věta říká, že tato mez je dosažitelná. Pro blokové kódy lze pravděpodobnost chyby libovolně snížit a přitom se libovolně přiblížit přenosové rychlosti dané kapacitou informačního kanálu. Pro obecný informační kanál je konstrukce založena na typické množině z Tvrzení 47. Je-li $P_A = (P_A(a))_{a\in A}$ pravděpodobnostní rozdělení na abecedě A, je jeho typická množina

$$\mathcal{M}^n_\varepsilon(P_A) \quad = \quad \left\{ x \in A^n : \ \left| \mathcal{H}(P_A) + \frac{1}{n} \log \prod_{i=0}^{n-1} P_A(x_i) \right| < \varepsilon \right\}$$

Definice 52 *Společná typická množina rozložení $P = (P(a,b))_{a\in A, b\in B}$ na $A \times B$ s marginálními rozloženími P_A na A a P_B na B je*

$$\overline{\mathcal{M}^n_\varepsilon}(P) = (\mathcal{M}^n_\varepsilon(P_A) \times \mathcal{M}^n_\varepsilon(P_B)) \ \cap \ \mathcal{M}^n_\varepsilon(P).$$

Společná typická množina je částí $A^n \times B^n$. Na obrázku 33 je společná typická množina $\mathcal{M}^5_{0.05}$ pro asymetrický binární kanál s parametry $r = 0.19$ a $s = 0.21$. Optimální vstupní a výstupní rozložení jsou $P_A = (0.503, 0.497)$, $P_B = (0.512, 0.488)$. Pro tyto parametry jsou individuální typické množiny $\mathcal{M}^5_{0.05}(P_A) = A^5$, $\mathcal{M}^5_{0.05}(P_B) = B^5$. Body typické množiny $\overline{\mathcal{M}^n_\varepsilon}(P_{A,B})$ jsou vyznačeny čtverečky. Zvolme na vstupu 3-separovanou množinu kódových slov $K = \{00000, 00111, 11011, 11100\}$ vyznačenou šipkami. Pak pro každé $y \in B^n$ existuje nejvýše jedno $x = f(y) \in K$, pro které $(f(y), y) \in \overline{\mathcal{M}^5_\varepsilon}(P_{A,B})$. Tímto způsobem se sestrojuje odhad $f : B^n \to A^n$.



Obrázek 33: Typická množina

7.4 Rychlost přenosu

Tvrzení 138 *Nechť* $(X, Y) = ((X_i, Y_i) : \Omega \to A \times B)_{i<n}$ *je posloupnost nezávislých náhodných veličin s rozložením* P *na* $A \times B$ *a nechť* $(\widetilde{X}, \widetilde{Y}) = ((\widetilde{X}_i, \widetilde{Y}_i) : \Omega \to A \times B)_{i<n}$ *je posloupnost nezávislých náhodných veličin s rozložením* $\mathbb{P}[\widetilde{X}_i = a, \widetilde{Y}_i = b] = P_A(a)P_B(b)$. *Pak pro každé* $\varepsilon, \delta > 0$ *existuje* $n_{\delta,\varepsilon}$ *takové, že pro všechna* $n > n_{\delta,\varepsilon}$ *platí*

(1) $\mathbb{P}[(X, Y) \in \overline{\mathcal{M}^n_\varepsilon}(P)] > 1 - \delta$

(2) $(1 - \delta)2^{n(\mathcal{H}(X_0,Y_0)-\varepsilon)} \leq |\overline{\mathcal{M}^n_\varepsilon}(P)| \leq 2^{n(\mathcal{H}(X_0,Y_0)+\varepsilon)}$

(3) $(1 - \delta)2^{-n(\mathcal{I}(X_0:Y_0)+3\varepsilon)} \leq \mathbb{P}[(\widetilde{X}, \widetilde{Y}) \in \overline{\mathcal{M}^n_\varepsilon}(P)] \leq 2^{-n(\mathcal{I}(X_0:Y_0)-3\varepsilon)}$

Důkaz: (1) Podle Věty 47 o rovnoměrném rozložení pro všechna dosti velká n platí

$$\mathbb{P}[X \in \mathcal{M}^n_\varepsilon(P_A)] > 1 - \tfrac{\delta}{3},\ \ \mathbb{P}[Y \in \mathcal{M}^n_\varepsilon(P_B)] > 1 - \tfrac{\delta}{3},\ \ \mathbb{P}[(X, Y) \in \mathcal{M}^n_\varepsilon(P)] > 1 - \tfrac{\delta}{3}$$

$$\mathbb{P}[(X, Y) \notin \overline{\mathcal{M}^n_\varepsilon}(P)] \leq \mathbb{P}[X \notin \mathcal{M}^n_\varepsilon(P_A)] + \mathbb{P}[Y \notin \mathcal{M}^n_\varepsilon(P_B)] + \mathbb{P}[(X, Y) \notin \mathcal{M}^n_\varepsilon(P)] \leq \delta.$$

(2a) Podle Tvrzení 47 je $|\overline{\mathcal{M}^n_\varepsilon}(P)| \leq |\mathcal{M}^n_\varepsilon(P)| \leq 2^{n(\mathcal{H}(X_0,Y_0)+\varepsilon)}$.
(2b) Naopak je-li $\mathbb{P}[(X, Y) \in \overline{\mathcal{M}^n_\varepsilon}(P)] > 1 - \delta$, je

$$(1 - \delta) \quad \leq \quad \sum_{(x,y)\in\overline{\mathcal{M}^n_\varepsilon}(P)} P^n(x, y) \leq |\overline{\mathcal{M}^n_\varepsilon}(P)| \cdot 2^{-n(\mathcal{H}(X_0,Y_0)-\varepsilon)}$$

$$(3)\ \ \mathbb{P}[(\widetilde{X}, \widetilde{Y}) \in \overline{\mathcal{M}^n_\varepsilon}(P)] \quad = \quad \sum_{(x,y)\in\overline{\mathcal{M}^n_\varepsilon}(P)} P^n_A(x)P^n_B(y) \leq |\overline{M^n_\varepsilon}(P)| \cdot 2^{-n(\mathcal{H}(X_0)+\mathcal{H}(Y_0)-2\varepsilon)}$$

$$\begin{aligned}
&\le 2^{n(\mathcal{H}(X_0,Y_0)-\mathcal{H}(X_0)-\mathcal{H}(Y_0)+3\varepsilon)} = 2^{-n(\mathcal{I}(X_0:Y_0)-3\varepsilon)} \\
\mathbb{P}[(\widetilde{X},\widetilde{Y}) \in \overline{\mathcal{M}^n_\varepsilon}(P)] &= \sum_{(x,y)\in\overline{\mathcal{M}^n_\varepsilon}(P)} P^n_A(x)P^n_B(y) \ge |\overline{\mathcal{M}^n_\varepsilon}(P)| \cdot 2^{-n(\mathcal{H}(X_0)+\mathcal{H}(Y_0)+2\varepsilon)} \\
&\ge (1-\delta)\cdot 2^{n(\mathcal{H}(X_0,Y_0)-\mathcal{H}(X_0)-\mathcal{H}(Y_0)-3\varepsilon)} \\
&= (1-\delta)\cdot 2^{-n(\mathcal{I}(X_0:Y_0)+3\varepsilon)}. \;\square
\end{aligned}$$

Věta 139 (Shannon) *Nechť $R : A \to B$ je informační kanál s kladnou kapacitou. Pak pro každé $\varepsilon > 0$ existuje $n > 0$, a náhodná veličina $X_n : \Omega \to A^n$, taková že platí*

$$\mathcal{E}_{X_n} < \varepsilon,\ \ \mathcal{H}(X_n)/n > C(R) - \varepsilon$$

Důkaz: Nechť $P_A \in \Delta(A)$ je rozdělení které maximalizuje $\mathcal{I}(X:Y)$ a položme $P(a,b) = P_A(a)R(a,b)$. Pro dané $\varepsilon > 0$ zvolme $n > n_{\varepsilon,\varepsilon}$ z Tvrzení 138, které navíc splňuje $\varepsilon + 2^{-n\varepsilon} < 2\varepsilon$ a $n > 1/\varepsilon$. Položme $m = \lceil 2^{n(C(R)-4\varepsilon)}\rceil$. Uvažujme kód $u \in (A^n)^m$, tj. posloupnost kódových slov $u_0,\ldots,u_{m-1}$ v A^n. Sestrojíme odhad $f_u : B^n \to A^n$. Pro $y \in B^n$ položme

$$f_u(y) = u_i,\ \ \text{kde}\ \ i = \min\{j < m :\ (u_j, y) \in \overline{\mathcal{M}^n_\varepsilon}(P)\}$$

Pokud takové $j < m$ neexistuje, zvolíme $f_u(y)$ libovolně. Označme Y_i náhodnou veličinu výstupu při vstupu $X_i = u_i$. Pro pravděpodobnost chyby při vstupu u_i dostáváme

$$\begin{aligned}
\mathcal{E}(f_u,u_i) &\le \mathbb{P}[(u_i,Y_i) \notin \overline{\mathcal{M}^n_\varepsilon}(P)\ \vee\ \exists j \ne i, (u_j,Y_i) \in \overline{\mathcal{M}^n_\varepsilon}(P)] \\
&\le \mathbb{P}[(u_i,Y_i) \notin \overline{\mathcal{M}^n_\varepsilon}(P)] + \sum_{j\ne i} \mathbb{P}[(u_j,Y_i) \in \overline{\mathcal{M}^n_\varepsilon}(P)]
\end{aligned}$$

Při rovnoměrném rozdělení na $\{u_0,\ldots,u_{m-1}\}$ je průměrná a maximální chyba

$$\overline{\mathcal{E}}(f_u) = \sum_{i=0}^{m-1} \mathcal{E}(f_u,u_i)/m,\ \ \mathcal{E}(f_u) = \max\{\mathcal{E}(f_u,u_i) :\ i < m\}$$

Místo kódu $u \in (A^n)^m$ uvažujme nyní náhodný kód, tj. náhodnou veličinu $X : \Omega \to (A^n)^m$. Vypočítáme očekávanou hodnotu $\mathbb{E}(\overline{\mathcal{E}}(f_X))$ chyby přenosu a z toho odvodíme že existuje kód který tuto očekávanou hodnotu nepřevyšuje. Náhodné veličiny $(X_{ij})_{i<m,j<n}$ jsou navzájem nezávislé s rozdělením P_A, tj. $\mathbb{P}[X_{ij} = a] = P_A(a)$. Náhodné veličiny $X_0,\ldots,X_{m-1} : \Omega \to A^n$ jsou také nezávislé. Nechť $Y_i : \Omega \to B^n$ jsou náhodné veličiny na výstupu kanálu R_n při vstupu X_i. Pro $i \ne j$ jsou X_i, Y_j nezávislé, takže

$$\mathbb{P}[X_i = u, Y_i = v] = P^n(u,v),\ \ \mathbb{P}[X_i = u, Y_j = v] = P^n_A(u)P^n_B(v)\ \ \text{pro}\ \ i \ne j$$

Počítejme nyní očekávanou pravděpodobnost chyby

$$\begin{aligned}
\mathbb{E}(\overline{\mathcal{E}}(f_X)) &= \sum_{u\in A^{mn}} \mathbb{P}[X=u]\cdot\overline{\mathcal{E}}(f_u) = \frac{1}{m}\sum_{u\in A^{mn}} \mathbb{P}[X=u]\sum_{i=0}^{m-1}\mathcal{E}(f_u,u_i) \\
&\le \frac{1}{m}\sum_{i=0}^{m-1}\sum_{u\in A^{mn}} \mathbb{P}[X=u]\cdot\mathbb{P}[(X_i,Y_i)\notin\overline{\mathcal{M}^n_\varepsilon}(P) \vee \exists j\ne i, (X_j,Y_i)\in\overline{\mathcal{M}^n_\varepsilon}(P) | X=u] \\
&= \frac{1}{m}\sum_{i=0}^{m-1}\mathbb{P}[(X_i,Y_i)\notin\overline{\mathcal{M}^n_\varepsilon}(P)\ \vee\ \exists j\ne i, (X_j,Y_i)\in\overline{\mathcal{M}^n_\varepsilon}(P)] \\
&\le \frac{1}{m}\sum_{i=0}^{m-1}\left(\mathbb{P}[(X_i,Y_i)\notin\overline{\mathcal{M}^n_\varepsilon}(P)] + \sum_{j\ne i}\mathbb{P}[(X_j,Y_i)\in\overline{\mathcal{M}^n_\varepsilon}(P)]\right) \\
&\le \frac{1}{m}\sum_{i=0}^{m-1}\left(\varepsilon + (m-1)\cdot 2^{-n(\mathcal{I}(X_0:Y_0)-3\varepsilon)}\right) = \varepsilon + (m-1)\cdot 2^{-n(\mathcal{I}(X_0:Y_0)-3\varepsilon)} \\
&\le \varepsilon + 2^{n(C(R)-4\varepsilon)}\cdot 2^{-n(C(R)-3\varepsilon)} = \varepsilon + 2^{-n\varepsilon} < 2\varepsilon
\end{aligned}$$

Existuje tedy kód $u \in (A^n)^m$ pro který $\overline{\mathcal{E}}(f_u) < 2\varepsilon$ a $m \geq 2^{n(C(R)-4\varepsilon)}$. Z kódu u vybereme jen ta slova, jejichž pravděpodobnost chyby nepřesahuje $2\overline{\mathcal{E}}(f_u)$. Tak dostaneme kód $v \in (A^n)^p$, kde

$$K_0 = \{v_0, \ldots, v_{p-1}\} = \{i < m : \mathcal{E}(f_u, u_i) \leq 2\overline{\mathcal{E}}(f_u)\}$$

Pak $p \geq m/2$ a pro odhad $f_v : B^n \to K_0$, a rovnoměrné rozložení X_n na K_0 platí

$$\mathcal{E}_{X_n} < \mathcal{E}(f_v) < 4\varepsilon, \quad \frac{\mathcal{H}(X_n)}{n} \geq \frac{\log m - 1}{n} > C(R) - 4\varepsilon - \frac{1}{n} > C(R) - 5\varepsilon. \ \square$$

Z Fanovy nerovnosti plyne, že pokud chceme dosáhnout malé chyby přenosu, nelze překročit rychlost přenosu danou kapacitou kanálu.

Věta 140 (Shannon) *Nechť $R : A \to B$ je informační kanál a $(X_n : \Omega \to A^n)_{n\geq 1}$ posloupnost náhodných veličin. Pak*

$$\lim_{n\to\infty} \mathcal{E}_{X_n} = 0 \implies \lim_{n\to\infty} \overline{\mathcal{E}_{X_n}} = 0 \implies \limsup_{n\to\infty} \frac{\mathcal{H}(X_n)}{n} \leq C(R)$$

Důkaz: Nechť Y_n je náhodná veličina na výstupu. Podle Fanovy nerovnosti platí

$$\mathcal{H}(X_n) \leq \mathcal{I}(X_n : Y_n) + \mathcal{H}(X_n|Y_n) \leq nC(R) + h(\overline{\mathcal{E}_{X_n}}) + \overline{\mathcal{E}_{X_n}} \cdot n \cdot \log|A|$$

Pro každé $\varepsilon > 0$ existuje $\delta > 0$ takové že pokud $\overline{\mathcal{E}_{X_n}} < \delta$, pak $h(\overline{\mathcal{E}_{X_n}}) + \overline{\mathcal{E}_{X_n}} \cdot \log|A| \leq \varepsilon$. Pro všechna n pro která je $\overline{\mathcal{E}_{X_n}} < \delta$ tedy platí $\frac{\mathcal{H}(X_n)}{n} \leq C(R) + \varepsilon$. Z toho plyne

$$\limsup_{n\to\infty} \frac{\mathcal{H}(X_n)}{n} \leq C(R) + \varepsilon.$$

Protože tato nerovnost platí pro každé $\varepsilon > 0$, je tím věta dokázána. $\square$

7.5 Lineární kódy

Důkaz Shannonovy věty je nekonstruktivní. Kód s malou chybou přenosu a optimální kapacitou lze získat náhodnými pokusy. Efektivní použití takového kódu je ale omezeno rychlostí dekódovací funkce $f_n : B^n \to K_n \subseteq A^n$. Není-li v K_n žádná symetrie, lze hodnotu $f_n(u)$ získat pouze prohlížením celé množiny K_n a to je časově náročné. Proto se hledají kódy, pro který má dekódovací funkce nízkou časovou složitost. Zde se používají algebraické metody. Abeceda A s q prvky se chápe jako množina zbytkových tříd $A = Z_q = \{0, 1, \ldots, q-1\}$. S operacemi sčítání a násobení modulo q je Z_q okruh. Pokud q je prvočíslo, je Z_q dokonce těleso. Množina slov A^n je pak vektorový prostor nad A dimenze n.

Definice 53 *Množina K je q-ární (n,m)-kód, pokud $K \subseteq \mathbb{Z}_q^n$ a $\#K = m$. Minimální průměr a poloměr kódu je*

$$d(K) = \min\{d(u,v) : u, v \in K, u \neq v\}, \quad t(K) = \left\lceil \frac{d(K)-1}{2} \right\rceil$$

Kód K je perfektní, pokud $\{B_{t(K)}(u) : u \in K\}$ je disjunktní rozklad množiny $\mathbb{Z}_q^n$.

Kód K detekuje nejvýše $d(K) - 1$ chyb a odstraňuje nejvýše $t(K)$ chyb. Množiny $B_{t(K)}(u)$, pro $u \in K$ jsou totiž disjunktní.

Tvrzení 141 *Je-li $K \subseteq A^n$ perfektní q-ární (n,m)-kód, pak $d(K) = 2 \cdot t(K) + 1$ a*

$$m \cdot \sum_{j=0}^{t(K)} \binom{n}{j} \cdot (q-1)^j = q^n$$

Důkaz: Pro každé $u \in A^n$ je $\#B_a(u) = \sum_{j=0}^{a} \binom{n}{j} \cdot (q-1)^j$. $\square$

Tvrzení 141 speciálně říká že q^n je dělitelné $|B_{t(K)}(u)|$. Tato podmínka je splněna pouze v jednom z následujících pěti případů

1. $q = 2$, $n = 2t + 1$, $m = 2$
2. $q \geq 2$, $t = 1$, $\exists r \geq 2$, $n = (q^r - 1)/(q - 1)$, $m = q^{n-r}$
3. $q = 3$, $t = 2$, $n = 11$, $m = 3^6$
4. $q = 2$, $t = 3$, $n = 23$, $m = 2^{12}$
5. $q = 2$, $t = 1$, $n = 90$, $m = 2^{78}$

První případ odpovídá opakovacímu kódu $K_n = \{0^n, 1^n\}$. Druhý případ nastává pro tzv. Hammingovy kódy. Třetí a čtvrtý případ nastává pro tzv. Golovayovy kódy. Pro pátý případ žádný perfektní kód neexistuje.

Definice 54 *Lineární (n, q^k)-kód je lineární podprostor $\mathbb{Z}_q^n$ dimenze k. Matice G typu $k \times n$ je generující matice K, tvoří-li její řádky bázi prostoru K. Matice H typu $(n-k) \times n$ je kontrolní matice K, tvoří-li její řádky bázi ortogonálního doplňku $K^\perp$ prostoru K.*

Příklad 21 *Opakovací kód $K = \{0000, 1111\}$ je binární $(4, 2^1)$-kód s maticemi*

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad d(K) = 4$$

Příklad 22 *Kód kontroly parity je binární $(4, 2^3)$-kód s maticemi*

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}, \quad d(K) = 2$$

Kód kontroly parity dostaneme z opakovacího kódu záměnou matice G a H.

Příklad 23 *Koktavý kód je binární $(4, 2^2)$-kód s maticemi*

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad d(K) = 2$$

Koktavý kód je samoduální, jeho generující a kontrolní matice jsou shodné.

Příklad 24 *Hammingův kód je binární $(7, 2^4)$-kód s maticemi*

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad d(K) = 3$$

Hammingův kód je perfektní s průměrem $d(K) = 3$. Jeho kontrolní matice je sestavena ze všech nenulových vektorů.

Věta 142 *Průměr lineárního kódu $d(K)$ je nejmenší číslo takové, že každých $d(K)$ sloupců kontrolní matice je lineárně závislých.*

Důkaz: Nechť $d = d(K)$ je průměr. Pak existují vektory $u, v \in K$, pro které $d(K) = d(u, v) = d(0, v - u)$, takže vektor $w = v - u \in K$ má alespoň $d(K)$ nenulových prvků $w_{i_1}, \ldots, w_{i_e}$, kde $e \geq d(K)$. Pro j-tý řádek matice H platí $H_{j,i_1} \cdot w_{i_1} + \cdots + H_{j,i_e} \cdot w_{i_e} = 0$

takže sloupce $i_1, \ldots, i_e$ matice H jsou lineárně závislé.
Naopak předpokládejme, že sloupce $i_1, \ldots, i_e$ matice H jsou lineárně závislé. Pak existuje vektor $w \in A^n$ pro který w_j je nenulové právě pro čísla $j = i_1, \ldots, i_e$ a $H \cdot w^T = 0$. Odtud $w \in K$ a $d(K) \leq d(0, w) = e$. $\square$

Pro $q \geq 2$ a $r \geq 2$ má lineární podprostor prostoru A^r dimenze 1 právě q prvků (včetně nulového). Rozdělíme-li $q^r - 1$ nenulových vektorů prostoru A^r do lineárních podprostorů, dostaneme v každém podprostoru $q-1$ prvků. Existuje tedy právě $(q^r-1)/(q-1)$ navzájem nezávislých (a tedy nenulových) vektorů v A^r. Kontrolní matice Hammingova kódu je sestavena právě z těchto lineárně nezávislých (sloupcových) vektorů. Je tedy $n = (q^r-1)/(q-1)$, $k = n-r$, $t(K) = 1$, $m = q^{n-r}$. Rovnost z Tvrzení 141 má tvar $m(1+n(q-1)) = q^{n-r} \cdot q^r = q^n$. Pro $q = 2$, $r = 3$ dostáváme kód z Příkladu 24. Pro $q = 2$, $r = 2$ dostáváme opakovací kód $K = \{000, 111\}$

7.6 Lineární dekódér

Nechť $K \subseteq A^n$ je lineární kód s generující maticí G a kontrolní maticí H. Jestliže $x \notin K$, je $H \cdot x^T \in A^{n-k}$ nenulový vektor, který se nazývá syndrom x (soubor příznaků). Definujme ekvivalenci

$$x \sim y \iff H \cdot x^T = H \cdot y^T$$

Třídy ekvivalance jsou afinní množiny tvaru $K_x = x+K = \{x+y : y \in K\}$ a $H(K_x) = H \cdot x^T$ nezávisí na výběru x. Například pro kód kontroly parity je

$$\begin{aligned} K_{000} &= \{000, 011, 101, 110\},\ H(K_{000}) = 0 \\ K_{001} &= \{001, 010, 101, 111\},\ H(K_{001}) = 1 \end{aligned}$$

Zvolme reprezentanta každé třídy ekvivalence, tj. výběrovou funkci $F : A^{n-k} \to A^n$ předpisem $F(0) = 000$, $F(1) = 001$. Dekódovací funkce je pak dána předpisem

$$f(u) = u - F(H \cdot u^T)$$

To znamená že od (výstupního) slova u odečteme reprezentanta jeho třídy ekvivalence. Při tomto dekódování se každé slovo zobrazí na první slovo sloupce, ve kterém se nachází. Pokud má lichou paritu, změní se jeho poslední bit.

Definice 55 *Nechť $K \subseteq \mathbb{Z}_q^n$ je lineární (n, q^k)-kód dimenze k. Výběrová funkce je zobrazení $F : \mathbb{Z}_q^{n-k} \to \mathbb{Z}_q^n$ pro kterou platí*

(1) $s \in A^{n-k} \implies H \cdot F(s)^T = s$

(2) $u \in A^n \ \&\ H \cdot u^T = s \implies d(u, 0) \geq d(F(s), 0)$

Tvrzení 143 *Nechť $F : A^{n-k} \to A^n$ je výběrová funkce lineárního kódu $K \subseteq A^n$ a pro $u \in A^n$ položme $f(u) = u - F(H \cdot u^T)$. Pak f je dekódovací funkce $f : A^n \to K$ a platí*

$$u \in A^n \ \&\ w \in K \implies d(u, f(u)) \leq d(u, w)$$

Důkaz: Pro $u \in A^n$ položme $s = H \cdot u^T$. Pak $H \cdot f(u)^T = s - H \cdot F(s)^T = 0$. Je-li $w \in K$, je $H \cdot (u - w)^T = s$ a

$$d(u, f(u)) \leq d(u - f(u), 0) = d(F(s), 0) \leq d(u - w, 0) = d(u, w)$$

Pro Hammingův kód je

$$
\begin{aligned}
K_0 &= \{0000000, 1000011, 0100101, 0010110, \ldots\}, \quad H(K_0) = 000 \\
K_1 &= \{1000000, 0000011, 1100101, 1010110, \ldots\}, \quad H(K_1) = 001 \\
K_2 &= \{0100000, 1100011, 0000101, 0110110, \ldots\}, \quad H(K_2) = 010 \\
K_3 &= \{0010000, 1010011, 0110101, 0000110, \ldots\}, \quad H(K_3) = 011 \\
K_4 &= \{0001000, 1001011, 0101101, 0011110, \ldots\}, \quad H(K_4) = 100 \\
K_5 &= \{0000100, 1000111, 0100001, 0010010, \ldots\}, \quad H(K_5) = 101 \\
K_6 &= \{0000010, 1000001, 0100111, 0010100, \ldots\}, \quad H(K_6) = 110 \\
K_7 &= \{0000001, 1000010, 0100100, 0010111, \ldots\}, \quad H(K_7) = 111
\end{aligned}
$$

Výběrová funkce je

$$F(s_0 s_1 s_2)_j = \begin{cases} 0 \text{ pro } j \neq 4s_0 + 2s_1 + s_2 \\ 1 \text{ pro } j = 4s_0 + 2s_1 + s_2 \end{cases}$$

Reference

[1] N. M. Abramson. *Information Theory and Coding.* McGraw-Hill, New York, 1963.

[2] Jiří Adámek. *Foundations of Coding.* Wiley-Interscience, 1991.

[3] G. Berkeley. *A Treatise Concerning the Principle of Human Knowledge.* Jacob Tonson, London, 1734.

[4] Patrick Billingsley. *Ergodic Theory and Information.* Wiley, New York, 1965.

[5] Leon Brillouin. *Science and Information Theory.* Academic Press, New York, 1962.

[6] G. J. Chaitin. *Algorithmic Information Theory.* Cambridge University Press, Cambridge, 1987.

[7] Kai Lai Chung. *Matkov Chains with Stationary Transition Probabilities.* Springer-Verlag, Berlin, 1960.

[8] T. M. Coven and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[9] I. Cziszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[10] Richard S. Ellis. *Entropy, Large Deviations and Statistical Mechanics.* Springer-Verlag, New York, 1985.

[11] Robert M. Gallager. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[12] R. M. Gray. *Entropy and Information Theory.* Springer-Verlag, Berlin, 1990.

[13] J.G. Kemeny and J.L. Snell. *Finite Markov chains.* Van Nostrand Reinhold, Princeton, 1960.

[14] A. I. Khinchin. *Mathematical Foundations of Information Theory.* Dover, 1957.

[15] Petr Kůrka. *Topological and Symbolic Dynamics.* Cours Spécialisés 11. Société Mathématique de France, Paris, 2003.

[16] S. Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[17] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Trans. Inform. Theory*, IT-22:75–81, 1976.

[18] L. A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14:1413–1416, 1973.

[19] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, Berlin, 1993.

[20] D. Lind and B. Marcus. *Symbolic Dynamicss and Coding.* Cambridge University Press, Cambridge, 1995.

[21] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-correcting Codes.* North-Holland, Amsterdam, 1977.

[22] Bruno Martin. *Codage, Cryptologie et Applications.* Presses Polytechniques et Universitaires Romandes, Lausanne, 2004.

[23] Nathaniel F. G. Martin, James W. England, and James K. Brooks. *Mathematical Theory of Entropy.* Addison-Wesley, Reading, 1981.

[24] D. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Trans. Inform. Theory*, 39:78–83, 1993.

[25] Alfréd Rényi. *Teorie pravděpodobnosti.* Academia, Praha, 1972.

[26] S. Roman. *Coding and Information Theory.* Springer-Verlag, Berlin, 1991.

[27] David Salomon. *Data Compression.* Springer-Verlag, Berlin, 2004.

[28] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[29] C. E. Shannon and W. W. Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1949.

[30] Paul C. Shields. *The Ergodic Theory of Discrete Sample Paths*, volume 13 of *Graduate Studies in Mathematics.* American Mathematical Society, 1996.

[31] Josef Štěpán. *Teorie pravděpodobnosti.* Academia, Praha, 1987.

[32] J. H. van Lint. *Introduction to Coding Theory.* North Holland, 1995.

[33] J. Wolfowitz. *Coding Theorems of Information Theory.* Prentice-Hall, Englewood Cliffs, 1978.