# Entropie v pravděpodobnostních dynamických systémech - Texty k přednášce

M Kupsa

4. února 2025

## Obsah

# 1 Úvod

Tato skripta se snaží pokrýt látku v předmětu “Entropie v pravděpodobnostních dynamických systémech”.

Hlavní ambicí předmětu je vyložit problematiku teorie informace aniž bychom se nechali svázat podmínkou nezávislosti. Klasická literatura na toto téma se rozvine do zajímavějších výsledků skrze slabý a silný zákon velkých čísel, čímž nevyhnutelně přijme omezující požadavek nezávislosti náhodných veličin, kterých se teorie týká. Takto vybudovaná teorie se pak dá vcelku dobře zobecnit na Markovské řetězce, ale tam bohužel možnosti teorie opřené o zákony velkých čísel končí.

Paralelně k tomuto přístupu byla vybudována teorie informace, či alespoň její část, v obecnějším kontextu stacionárních procesů, která se namísto zákona velkých čísel opírá o ergodickou větu.

Teorie informace v našem podání se tedy bude týkat stacionárních procesů, konkrétně podtřídy ergodických procesů. Zavedeme klíčový pojem entropie, informačního obsahu, informační hustoty. Vlastnosti těchto veličin nakonec využijeme k důkazu efektivity kompresních algoritmů. Tímto pokryjeme takzvanou "Teorii kódování zdroje". Druhou částí klasické teorie informace je teorie kódování kanálu. Touto se zde zabývat nebudeme.

Namísto toho zavedeme pojem entropie také pro pravděpodobnostní dynamické systémy a ukážeme, že se jedná o velmi užitečnou číselnou charakteristiku systémů, která je spolu s ergodickou a spektrální teorií klasickou součástí ergodické teorie.

Nejprve ovšem uvedeme následující motivaci k pojmu entropie a informace.

## 1.1 Entropie jako cena informace

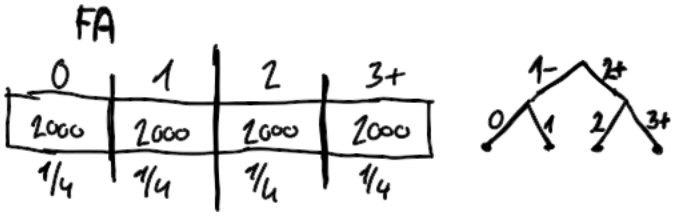
Mějme nejmenovanou instituci, která má seznam 8000 lidí (říkejme jim souhrnně populace) a ráda by zjistila, kolik kreditních karet má každý z nich. Zadá tento důležitý úkol raději hned dvěma různým firmám.

Firma FA přinese seznam 8000 lidí, ve kterém jsou v kolonce počtu karet zapsány cifry 0,1,2 a 3+, kde 3+ znamená tři a více bez přesnějšího rozlišení. Počet lidí příslušející každé kategorii je 2000, viz Obrázek 1. Firma FB přinese stejný seznam, ve kterém je zapsáno 1-,2,3 a 4+, kde 1- znamená 1 nebo žádná, a 4+ znamená 4 a více. Počet lidí pro příslušné počty karet jsou po řadě 4000, 2000, 1000 a 1000. Viz Obrázek 2.

Otázka je, který seznam je z hlediska informační hodnoty kvalitnější. Toto samozřejmě souvisí s konkrétním využitím, ale zjednodušme si nyní situaci a měřme to pouze přes „náklady“ na získání informace v podobě počtu otázek. V zájmu rovných podmínek, uvažujme pouze otázky, na které se dá odpovědět ANO a NE. Kolik otázek musela položit první a druhá firma? Firma FA se mohla například ptát, zda má člověk 2 a více karet, a dle získané odpovědi následnou otázkou rozlišit 0 nebo 1, případně 2 a 3+. Tato strategie se dá popsat pomocí stromu otázek, případně pomocí posloupnosti „řezů“ populací, viz Obrázek 1b, kde pořadí řezů populací je indikován délkou svislé čáry (nejdelší je první).

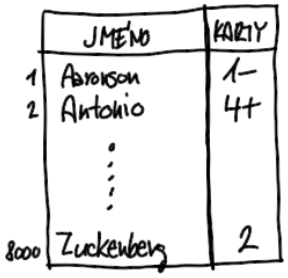
Tímto způsobem položí dohromady 16000 otázek, 2 otázky na člověka. Druhá firma může postupovat podobně, rozdělí si populaci první otázkou na dvě a dvě kategorie, a druhou otázkou se již dostane na jednu ze čtyř možností. Takto bude pokládat také 16000 otázek. Druhá firma má ale lepší možnost. Namísto strategie, která balancuje strom otázek z hlediska počtu možných odpovědí, může balancovat otázky z hlediska rozložení populace. Tento přístup vede k tomu, že první otázkou "1- vs. 2+"rozdělí populaci napůl. V případě odpovědi 1- se už nemusí dále ptát, v druhém případě opět rozdělí příslušnou část populace napůl otázkou „2 vs 3+“. V případě odpovědi „3+“, je pak třeba položit ještě třetí otázku „3 vs 4+“, viz Obrázek 2b. Takto se sice


(a) Seznam FA

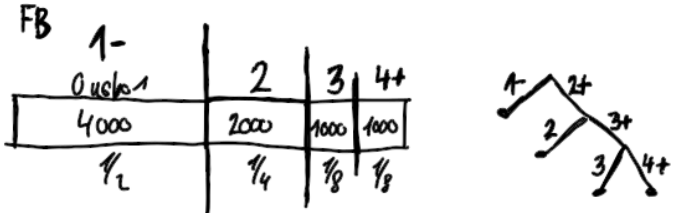



(b) Rozložení počtu kreditek, strom otázek

Obrázek 1: Průzkum dle firmy FA




(a) Seznam FB




(b) Rozložení počtu kreditek, strom otázek

Obrázek 2: Průzkum dle firmy FB

stane, že budeme pokládat některým lidem tři otázky, ale celkem bude položeno jen $4000+2000*2+2000*3 = 14000$ otázek, v průměru $7/4$ otázek na člověka. Firma FA tedy přinesla nákladnější informaci. Otázkou je, zda nákladnější znamená lepší.

Podívejme se nyní na věc pohledem klienta, neboli nejmenované instituce. Jak porovnat oba seznamy? Pokud by se jeden seznam dal zcela odvodit z druhého, jistě bychom prohlásili ten první za hodnotnější. To ovšem není tento případ, neboť seznamy se navzájem doplňují. Pokud je použijeme oba, jsme schopni u každého člověka říci, zda má 0,1,2,3 nebo 4+ kreditních karet. Takovéto spojení informací od obou firem do jednoho seznamu nabízí jiný způsob jak zjistit kvalitu informace. Konkrétně se můžeme ptát na dodatečné náklady, pokud by instituce chtěla doplnit seznam od firmy FA na sdruženou informaci. Firma FA by v takovém případě musela rozdělit kategorii 3+ na kategorie 3 a 4+. K tomu by stačilo položit $2000 * 1 = 2000$ otázek. Přepočteno na celkovou populaci to dává $1/4$ dodatečné otázky na člověka. Při doplňování seznamu od firmy FB na sdruženou informaci bychom potřebovali rozlišit kategorii 1− na kategorie 0 a 1. K tomu je třeba dodatečně položit $4000 * 1 = 4000$ otázek. V průměru $1/2$ otázky na člověka (vztaženo k celé populaci). I z tohoto pohledu je tedy informace od firmy FA kvalitnější.

Dejme nyní tento příklad do souvislosti s matematickými pojmy. Populaci označíme $\Omega$ a seznamy od firem FA a FB budeme chápat jako funkce $X : \Omega \to A$ a $Y : \Omega \to B$, kde $A = \{0, 1, 2, 3+\}$ a $B = \{1-, 2, 3, 4+\}$. Na množině $\Omega$ dále uvažujme pravděpodobnost $\mathbb{P}$, která každému člověku přiřadí nějakou váhu, v našem případě všem stejnou. Tím se funkce $X$ a $Y$ stanou náhodnými veličinami v širším slova smyslu. V teorii pravděpodobnosti se za náhodné veličiny standardně uvažují funkce z pravděpodobnostního prostoru do reálných čísel, což umožňuje správně definovat střední hodnotu, rozptyl atd. Pro teorii informace je přirozenější zabývat se zobrazeními do konečné, či spočetné množiny, kdy nepředpokládáme u možných hodnot žádnou

další strukturu, či zavedené operace. Proto není problém, že je hodnotou například barva, typ vozu, či označení„3+“, jako v našem příkladě. Pro naše dvě náhodné veličiny pak definujeme entropii $H(X)$ (resp. $H(Y)$), sdruženou entropii $H(X,Y)$ a podmíněnou entropii $H(X|Y)$ (resp. $H(Y|X)$) pomocí vzorců

$$\begin{aligned} H(X) &= -\sum_{a\in A} \mathbb{P}(X=a)\log \mathbb{P}(X=a), \\ H(X,Y) &= -\sum_{a\in A, b\in B} \mathbb{P}(X=a,Y=b)\log \mathbb{P}(X=a,Y=b), \\ H(X|Y) &= -\sum_{a\in A, b\in B} \mathbb{P}(X=a,Y=b)\log \mathbb{P}(X=a|Y=b), \end{aligned}$$

kde nedefinované členy sum ($0\log 0$, $0\log\frac{0}{0}$) ignorujeme, t.j. považujeme je za nulu. Základ logaritmu je fixní a obvykle ho volíme rovný 2. Jednoduchým výpočtem pak lze zjistit, že

$$H(X)=2, \qquad H(Y)=\frac{7}{4}, \qquad H(X,Y)=\frac{9}{4}, \qquad H(X|Y)=\frac{1}{2}, \qquad H(Y|X)=\frac{1}{4}.$$

Veličiny $H(X)$ a $H(Y)$ číselně odpovídají průměrnému počtu otázek na člověk, které musí položit firma FA či FB. Platí dokonce, že ve vzorci průměrované hodnoty $-\log\mathbb{P}(X=a)$ odpovídají přesně počtu otázek položených člověku z kategorie „$a$“ v ideální strategii pro firmu FA, podobně $-\log\mathbb{P}(Y=b)$ odpovídá počtu otázek pro člověka z kategorie „$b$“ při ideální strategii pro firmu FB. Podobně, $-\log\mathbb{P}(X=a\,|\,Y=b)$ zde odpovídá počtu dodatečných otázek pro člověka z kategorie $a$, víme-li od firmy FB, že člověk patří do kategorie „$b$“, z čehož pak plyne, že $H(X|Y)$ je rovno průměrnému počtu doplňujících otázek na člověka při znalosti seznamu od firmy FB.

Entropii se tedy dá rozumět jako průměrné kvantitě informace, kterou získáme, jsme-li schopni rozdělit jednolitou populaci do určitých skupin, například dle počtu kreditních karet. Jde tedy o kvantifikaci naší schopnosti rozlišovat. podobně se dá interpretovat sdružená entropie, kde rozlišujeme do jemnějších kategorií. Podmíněná entropie $H(X|Y)$ pak je kvantifikace schopnosti rozlišit v populaci kategorie veličiny $X$, pokud už umíme rozlišovat do kategorií veličiny $Y$.

Na závěr ještě zmiňme, že rovnost průměrného počtu otázek v různých scénářích s příslušnou entropií bývá v obecném případě jen přibližná. Takto pěkně příklad vyšel díky tomu, že všechny uvažované pravděpodobnosti, včetně těch podmíněných, byly celočíselnými mocninami čísla, které jsme zvolili jako základ pro logaritmus ve vzorcích pro entropii.

# 2 Informační obsah a entropie náhodné veličiny a příslušného rozkladu

Informační obsah, potažmo entropie, se dá definovat pro několik druhů objektů, které k sobě mají úzkou vazbu a jinými prostředky popisují stejný koncept. Nejčastějším přístupem je definovat tyto pojmy pro diskrétní náhodné veličiny. S touto definicí budeme pracovat velkou část přednášek také my. Větší důraz ovšem budeme klást na pojem rozkladu pravděpodobnostního prostoru, neboť ten dle našeho soudu poskytuje lepší představu o tom, jak teorie informace vnímá náhodné veličiny a pojem informace. Konkrétně vnímáme entropii jako kvantifikaci toho, jak daná náhodná veličina dokáže rozlišovat mezi elementy pravděpodobnostního prostoru, na kterém je definovaná, t.j. jak „jemný“ rozklad indukuje.

Buď $(\Omega,\mathcal{F},\mathbb{P})$ pravděpodobnostní prostor, tedy $\Omega$ je množina, $\mathcal{F}$ je $\sigma$-algebra podmnožin a $\mathbb{P}$ je pravděpodobnostní míra na $\mathcal{F}$. *Rozkladem pravděpodobnostního prostoru* $\Omega$ je každá spočetná,

či konečná množina $\mathcal{R} \subset \mathcal{F}$, která sestává ze vzájemně disjunktních množin jejichž sjednocení má pravděpodobnost 1. *Entropie rozkladu* $\mathcal{R}$ je definovaná následovně:

$$H(\mathcal{R}) = \sum_{\substack{R \in \mathcal{R} \\ \mathbb{P}(R)>0}} \mathbb{P}(R)\,(-\log \mathbb{P}(R))\,.$$

Tato suma sestává pouze z nezáporných sčítanců, proto je vždy definovaná. Pro spočetné (nekonečné) rozklady může být nekonečná, pro konečné rozklady je konečná. Požadavek $\mathbb{P}(R) > 0$ v sumě bývá v literatuře často vypouštěn, výraz $0 \log 0$ se interpretuje jako 0. Tento přístup ale může v některých případech vést k problémům, proto je třeba se mu buď zcela vyhnout, nebo si pohlídat jeho limity. Pro příjemnější zápis zavedeme pojem *nosiče rozkladu* (značíme $s$ z anglického „support“):

$$s(\mathcal{R}) = \{R \in \mathcal{R} \mid \mathbb{P}(R) > 0\}.$$

Potom tedy

$$H(\mathcal{R}) = \sum_{R \in s(\mathcal{R})} \mathbb{P}(R)\,(-\log \mathbb{P}(R))\,.$$

Každá diskrétní náhodná veličina $X$ definovaná na $\Omega$ s hodnotami v konečné či spočetné množině $A$ indukuje rozklad $\Omega$ (*jádro* zobrazení $X$):

$$\mathcal{R}_X = \{X^{-1}\{a\} \mid a \in A\}$$

a také pravděpodobnost $P_X$ na množině $A$, kde

$$P_X(a) := \mathbb{P}(X = a) = \mathbb{P}(X^{-1}(a)).$$

*Entropie náhodné veličiny* je definována jako

$$H(X) = \sum_{\substack{R \in \mathcal{R} \\ \mathbb{P}(R)>0}} P_X(a)(-\log P_X(a)).$$

Přímo z definic plyne, že $H(X) = H(\mathcal{R}_X)$.

Na druhou stranu, každý konečný nebo spočetný rozklad $\mathcal{R}$ je možné oindexovat konečnou nebo spočetnou množinou $A$, tak že

$$\mathcal{R} = \{R_a \mid a \in A\},$$

a definovat náhodnou veličinu $X$ předpisem $X(\omega) = a$, pokud $\omega \in R_a$. Potom $\mathcal{R}_X = \mathcal{R}$ a $H(X) = H(\mathcal{R})$. Každý rozklad tedy odpovídá nějaké náhodné veličině, ovšem z konstrukce je zřejmé, že takových veličin je nekonečně mnoho vzhledem k absolutní libovůli výběru indexové množiny. Dokonce i při stejné množině indexů stačí permutace při indexování k vytvoření různých náhodných veličin pro stejný rozklad. Zároveň se ukazuje, že rozklad určuje entropii, t.j. veličiny, které indukují stejný rozklad, mají stejnou entropii.

Entropii je vhodné chápat také jako střední hodnotu *informačního obsahu* $\mathcal{I}_\mathcal{R} : \Omega \to \mathbb{R}$. Informační obsah nejprve definujme pro množinu $M \in \mathcal{F}$ kladné pravděpodobnosti, předpisem $\mathcal{I}(M) = -\log \mathbb{P}(M)$. Pro rozklad $\mathcal{R}$ pak užijeme definici:

$$\mathcal{I}_\mathcal{R}(\omega) := \mathcal{I}(\mathcal{R}(\omega)) = -\log(\mathbb{P}(\mathcal{R}(\omega))),$$

kde $\mathcal{R}(\omega)$ je ta množina z $\mathcal{R}$, ve kterém je $\omega$. Pro dobře definovaný informační obsah potřebujeme aby, pravděpodobnost takové množiny byla kladná. To je splněno pro všechna $\omega$ ze sjednocení

množin z nosiče rozkladu, což je množina plné míry. Tedy informační obsah je dobře definovaný skoro všude. Funkce je konstantní na jednotlivých množinách z nosiče a jistě platí

$$\mathbb{E}(\mathcal{I}_{\mathcal{R}}) = \sum_{R \in s(\mathcal{R})} -\log(R)\mathbb{P}(R) = H(\mathcal{R}).$$

Informační obsah pro náhodnou veličinu $X$ pak ztotožníme z informačním obsahem příslušného rozkladu, t.j. $\mathcal{I}_X := \mathcal{I}_{\mathcal{R}_X}$. Platí tedy také, že $\mathbb{E}(\mathcal{I}_X) = H(X)$.

Pro dva rozklady $\mathcal{R}$ a $\mathcal{R}'$ na $\Omega$ zavedeme jejich *sdružený rozklad* $\mathcal{R} \vee \mathcal{R}'$ předpisem

$$\mathcal{R} \vee \mathcal{R}' = \{R \cap R' \mid R \in \mathcal{R}, R' \in \mathcal{R}'\}.$$

Je snadno vidět, že je to opět rozklad $\Omega$, který je zjemněním obou uvažovaných rozkladů. Informačnímu obsahu $\mathcal{I}_{\mathcal{R}\vee\mathcal{R}'}$ a entropii $H(\mathcal{R}\vee\mathcal{R}')$ budeme říkat sdružený informační obsah respektive sdružená entropie pro rozklady $\mathcal{R}$ a $\mathcal{R}'$.

Tato operace dobře odpovídá „sdružené“ náhodné veličině následujícím způsobem. Pokud máme dvě náhodné veličiny $X : \Omega \to A$ a $Y : \Omega \to B$, pak je přirozeně definovaná sdružená náhodná veličina $(X, Y) : \Omega \to A \times B$. Prvky $\mathcal{R}_{(X,Y)}$ jsou podle definice množiny tvaru

$$R_{a,b} = \{\omega \mid (X, Y)(\omega) = (a, b)\} = \{\omega \mid X(\omega) = a, Y(\omega) = b\}.$$

Z toho přímo plyne, že

$$\mathcal{R}_{(X,Y)} = \mathcal{R}_X \vee \mathcal{R}_Y,$$

neboť

$$\mathcal{R}_{(X,Y)} = \{(X, Y)^{-1}(a, b) \mid (a, b) \in A \times B\} = \{X^{-1}(a) \cap Y^{-1}(b) \mid a \in A, b \in B\}.$$

Neboli,

$$H(X, Y) = H(\mathcal{R}_X \vee \mathcal{R}_Y).$$

Pro rozklady z příkladu o kreditních kartách dostáváme, že

$$\begin{aligned}\mathcal{R}_{(X,Y)} &= \{X^{-1}(a) \cap X^{-1}(b) \mid a \in \{0, 1, 2, 3+\}, b \in \{1-, 2, 3, 4+\}\} \\ &= \{X^{-1}(0), X^{-1}(1), X^{-1}(2), Y^{-1}(3), Y^{-1}(4+), \emptyset\}.\end{aligned}$$

Poslední rovnost platí díky tomu, že mnoho množin tvaru $X^{-1}(a) \cap Y^{-1}(b)$ je prázdných.

Pro dvojici rozkladů můžeme také definovat podmíněný informační obsah $\mathcal{I}_{\mathcal{R}|\mathcal{R}'}$ a podmíněnou entropii $H(\mathcal{R}|\mathcal{R}')$:

$$\mathcal{I}_{\mathcal{R}|\mathcal{R}'}(\omega) = -\log(\mathbb{P}(\mathcal{R}(\omega)|\mathcal{R}'(\omega))), \qquad H(\mathcal{R}|\mathcal{R}') = \mathbb{E}(\mathcal{I}_{\mathcal{R}|\mathcal{R}'}).$$

Uvědomme si, že $\mathcal{I}_{\mathcal{R}|\mathcal{R}'}$ je dobře definován právě na množinách $R \cap R'$, $R \in \mathcal{R}$, $R' \in \mathcal{R}'$, které mají nenulovou pravděpodobnost. To jsou právě ty z nosiče $s(\mathcal{R} \vee \mathcal{R}')$. Na těchto množinách je konstantní a nabývá hodnoty $-\log \frac{\mathbb{P}(R\cap R')}{\mathbb{P}(R')}$.

Pro náhodné veličiny pak můžeme definovat podobně

$$\mathcal{I}_{X|Y} := \mathcal{I}_{\mathcal{R}_X|\mathcal{R}_{X'}}, \qquad H(X|Y) = \mathbb{E}(\mathcal{I}_{X|Y}).$$

Poslední definicí je pak vzájemný informační obsah, resp. vzájemná informace, definovaná předpisem

$$\mathcal{I}_{\mathcal{R}:\mathcal{R}'}(\omega) = \log \frac{\mathbb{P}(\mathcal{R}(\omega) \cap \mathcal{R}'(\omega))}{\mathbb{P}(\mathcal{R}(\omega) \cdot \mathbb{P}(\mathcal{R}'(\omega))}, \qquad H(\mathcal{R} : \mathcal{R}') = \mathbb{E}(\mathcal{I}_{\mathcal{R}:\mathcal{R}'}).$$

a

$$\mathcal{I}_{X:Y} := \mathcal{I}_{\mathcal{R}_X:\mathcal{R}_{X'}}, \qquad H(X:Y) = \mathbb{E}(\mathcal{I}_{X:Y}).$$

Ačkoliv jsme vysvětlili, že pojem informace je svázán primárně s rozkladem prostoru, uvedeme zde i úplné znění vzorců pro náhodné veličiny. Pro tyto pojmy pak zavedeme základní vlastnosti. Děláme tak zejména proto, že mnoho pojmů v teorii pravděpodobnosti, jako nezávislost atp., jsou definované právě pro veličiny. Laskavý čtenář by jistě zjistil, že jsou to opět pojmy primárně svázané s rozklady.

# 3 Vlastnosti informačního obsahu a entropie náhodné veličiny a příslušného rozkladu

Pro náhodnou veličinu s hodnotami v konečné (resp. spočetné) abecedě $A$ zavedeme pojem nosiče veličiny $X$ a rozdělení $P_X$,

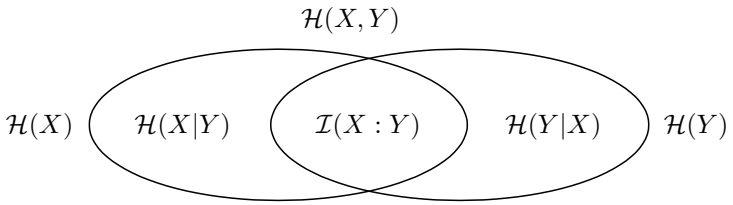
$$s(X) = \{\omega \in \Omega \mid P_X(X(\omega)) > 0\}, \qquad s(P_X) = \{a \in A \mid P_X(a) > 0\}.$$

Platí $\mathbb{P}(s(X)) = 1$ a $P_X(s(P_X)) = 1$.

Informační obsah i entropie, včetně podmíněné, jsou z definice nezáporné. Z definice plynou základní rovnosti:

$$\begin{aligned} \mathcal{I}_{X,Y} &= \mathcal{I}_X + \mathcal{I}_{Y|X} = \mathcal{I}_X + \mathcal{I}_Y - \mathcal{I}_{X:Y} && ,s.j. \\ H(X,Y) &= H(X) + H(Y|X) = H(X) + H(Y) - H(X:Y) \end{aligned}$$

Tyto vztahy dohromady tvoří užitečný diagram na Obrázku 3. Z něj lze názorně získávat další




Obrázek 3: Podmíněná a vzájemná entropie

rovnosti, např. následující alternativní definici vzájemného informačního obsahu a vzájemné entropie:

$$\begin{aligned} \mathcal{I}_{X:Y} &= \mathcal{I}_X + \mathcal{I}_Y - \mathcal{I}_{X,Y}, \\ H(X:Y) &= H(X) + H(Y) - H(X,Y)\,. \end{aligned}$$

Pozor, vzájemný informační obsah téměř vždy nabývá i záporné hodnoty (s kladnou pravděpodobností). Ukážeme ovšem, že jeho střední hodnota, tedy vzájemná informace, je nezáporná. Vzorce uvedené do této chvíle budeme obvykle nazývat součtovými vzorci pro entropii, resp. informační obsah. Důležité budou také jejich podmíněné verze:

$$\begin{aligned} \mathcal{I}_{X,Y|Z} &= \mathcal{I}_{X|Z} + \mathcal{I}_{Y|X,Z}, \quad s.j. \\ H(X,Y|Z) &= H(X|Z) + H(Y|X,Z). \end{aligned}$$

Přičtením $\mathcal{I}_Z$ k oběma stranám rovnice totiž dostáváme nepodmíněný součtový vzorec pro $Y$ a $(X, Z)$:

$$\mathcal{I}_{X,Y,Z} = \mathcal{I}_{X,Z} + \mathcal{I}_{Y|X,Z}, \quad s.j.$$

Opakovaným použitím těchto pravidel pak dostáváme následující řetězová pravidla.

**Tvrzení 1.** *Pro posloupnost náhodných veličin $X_0, \dots X_{n-1}$, $Z$ platí*

$$\begin{aligned}
\mathcal{I}_{X_{[0,n)}} &= \sum_{i=0}^{n-1} \mathcal{I}_{X_i|X_{[0,i)}} \ s.j. \\
H(X_{[0,n)}) &= \sum_{i=0}^{n-1} H(X_i|X_{[0,i)}) \\
\mathcal{I}_{X_{[0,n)}|Z} &= \sum_{i=0}^{n-1} \mathcal{I}_{X_i|X_{[0,i)},Z} \ s.j. \\
H(X_{[0,n)}|Z) &= \sum_{i=0}^{n-1} H(X_i|X_{[0,i)}, Z).
\end{aligned}$$

Pro množiny $M, N \in \mathcal{F}$ kladné míry platí

$$\begin{aligned}
\mathcal{I}(M) \geq \mathcal{I}(N), &\qquad \text{pro } M \subset N. \\
\mathcal{I}(M \cap N) = \mathcal{I}(M) + \mathcal{I}(N) &\qquad \text{právě když } M \perp N.
\end{aligned}$$

Zároveň platí, že nulový informační obsah a nulovou entropii má pouze triviální náhodná veličina, tedy taková, která nabývá jediné hodnoty s pravděpodobností 1. Pro takovou $Triv$ a libovolnou diskrétní náhodnou veličinu $X$ platí:

$$\mathcal{I}_{X|Triv} = \mathcal{I}_X, \quad H(X|Triv) = H(X).$$

## 3.1 Determinovanost

**Definice 1.** *Pokud pro náhodné veličiny $X : \Omega \to A$ a $Y : \Omega \to A'$ existuje funkce $f : s(P_X) \to s(P_Y)$ taková, že platí $Y = f(X)$ s.j., řekneme, že $X$ determinuje $Y$ ($Y$ je funkcí $X$.) Pokud je $f$ bijekcí mezi $s(P_X)$ a $s(P_y)$, pak říkáme, že $X$ je transformací $Y$. V prvním případě píšeme $X \vdash Y$, v druhém $X \sim Y$.*

v dalších úvahách je dobré chápat $s(P_{X,Y}) \subset A \times B$ jako relaci. Právě ona se stane funkcí $f$ pro případ determinované veličiny (jiná volba není možná).

**Lemma 1.** *Mějme náhodné veličiny $X : \Omega \to A$ a $Y : \Omega \to A'$. Potom pro relaci $s(P_{X,Y})$ platí*

$$\begin{aligned}
\pi_1(s(P_{X,Y})) &:= \{a \in A \mid \exists b \in B, (a, b) \in s(P_{X,Y})\} = s(P_X) \\
\pi_2(s(P_{X,Y})) &:= \{b \in B \mid \exists a \in A, (a, b) \in s(P_{X,Y})\} = s(P_Y).
\end{aligned}$$

*Pokud existuje funkce $f : s(P_X) \to s(P_Y)$ taková, že platí $Y = f(X)$ s.j., pak $f = s(P_{X,Y})$.*

*Důkaz.* Připomeňme, že $P_{X,Y}(a,b) = \mathbb{P}(X=a, Y=b)$. Jistě platí $\mathbb{P}(X=a) \geq \mathbb{P}(X=a, Y=b)$ a $\mathbb{P}(Y=b) \geq \mathbb{P}(X=a, Y=b)$ (jedná se o porovnání pravděpodobností množin, které jsou v relaci inkluze). Tedy z $(a,b) \in s(P_{X,Y})$, plyne $a \in s(P_X)$ a $b \in s(P_Y)$, t.j. $s(P_{X,Y}) \subset s(P_X) \times s(P_Y)$. Neboli $\pi_1(s(P_{X,Y})) \subset s(P_X)$, $\pi_2(s(P_{X,Y})) \subset s(P_Y)$.

Opačné nerovnosti dokážeme následovně. Označme $\Omega'$ množinu těch $\omega$, kde je $Y(\omega)$ dobře definované. Jistě $\mathbb{P}(\Omega') = 1$. Mějme nyní $a \in s(P_X)$. Potom

$$0 < P_X(a) = \mathbb{P}(X=a) = \mathbb{P}((X=a) \cap \Omega') = \sum_{b \in B} \mathbb{P}(X=a, Y=b).$$

Tedy alespoň jeden ze sčítanců musí být kladný, t.j. existuje $b \in B$, takové, že $P_{X,Y}(a,b) > 0$. Dostáváme, že $a \in \pi_1(s(P_{X,Y}))$. Jelikož jsme $a \in s(P_X)$ volili libovolně,máme $s(P_X) \subset \pi_1(s(P_{X,Y}))$. Obdobně dokážeme $s(P_Y) \subset \pi_2(s(P_{X,Y}))$.

Zbývá ukázat poslední část tvrzení, kde předpokládáme, že existuje $f : s(P_X) \to s(P_Y)$ taková, že platí $Y = f(X)$ s.j.. Označme $\Omega'$ množinu všech $\omega$, takových, že $f(X(\omega))$ a $Y(\omega)$ jsou dobře definované a $Y(\omega) = f(X(\omega))$. Jistě $\mathbb{P}(\Omega') = 1$. Musí tedy platit, že

$$1 = \mathbb{P}(\Omega') = \sum_{a \in s(P_X)} \mathbb{P}(X=a, Y=f(a)) = \sum_{a,b \in s(P_{X,Y})} \mathbb{P}(X=a, Y=b).$$

Z toho plyne, že pro $(a,b) \notin f$ je $P_{X,Y}(a,b) = 0$. Tedy $s(P_{X,Y}) \subset f$. Nechť $(a,b) \in f$. Tedy $a \in s(P_X)$ a $b = f(a)$. Dále

$$0 < \mathbb{P}(X=a) = \mathbb{P}(X=a, Y=f(a)) + \sum_{b' \in B, b' \neq f(a)} \mathbb{P}(X=a, Y=b')) = \mathbb{P}(X=a, Y=f(a)).$$

Tedy $(a,b) \in s(P_{X,Y})$. Dokázali jsme tedy i opačnou nerovnost, $f \in s(P_{X,Y})$. □

**Lemma 2.** *Diskrétní náhodné veličiny $X$ a $Y$ jsou vzájemnou transformací, pokud jedna determinuje druhou. Relace $X \sim Y$ je tedy symetrická.*

Poznamenejme, že symetričnost se dá odvodit přímo z definice. Ovšem korektní důkaz musí správně pracovat s množinami míry nula.

*Důkaz.* Z definice plyne, že z $X \sim Y$ plyne $X \vdash Y$ a $Y \vdash X$.

Pokud naopak $X \vdash Y$ a $Y \vdash X$, existuje $f : s(P_X) \to s(P_Y)$ a $g : s(P_Y) \to s(P_X)$ takové, že $f(X) = Y$ s.j. a $g(Y) = X$ s.j. Z předchozího lemmatu plyne, že $f = s(P_{X,Y})$ a $g = s(P_{Y,X})$. Ovšem $s(P_{X,Y})$ a $s(P_{Y,X})$ jsou z definice vzájemně inverzní relace. Zároveň víme, že $\pi_1(s(P_{X,Y})) = s(P_X)$ a $\pi_2(s(P_{X,Y})) = s(P_Y)$. Tedy $f$ je bijekce mezi $s(P_X)$ a $s(P_Y)$. Proto $X \sim Y$. □

**Lemma 3.** *Pro diskrétní náhodné veličiny $X, Y$, kde $X \vdash Y$, platí*

$$\mathcal{I}_Y \leq \mathcal{I}_X \text{ s.j.}, \qquad H(Y) \leq H(X).$$

*Důkaz.* Nechť existuje $f : s(X) \to s(Y)$ takové, že $f(X) = Y$ s.j. Označme $\Omega' \subset \Omega$ množinu bodů, pro které nastane rovnost. Potom pro $\omega \in \Omega' \cap s(X) \cap s(Y)$, $a = X(\omega)$, $b = f(a)$ platí inkluze

$$\mathcal{R}_X(\omega) = X^{-1}(a) \subset X^{-1}(f^{-1}(b)) = \mathcal{R}_Y(\omega).$$

Z toho plyne $\mathcal{I}_Y(\omega) \leq \mathcal{I}_X(\omega)$. □

V dalším tvrzení ukážeme nerovnice, které jsou s determinovaností ekvivalentní. Pozor, nerovnice z předchozího lemmatu takové nejsou.

**Tvrzení 2.** *Pro diskrétní náhodné veličiny $X, Y$, jsou následující podmínky ekvivalentní:*

1. $X \vdash Y$
2. $\mathcal{I}_{Y|X} = 0$ *s.j.*
3. $\mathcal{I}_{Y|X} \leq 0$ *s.j.*
4. $\mathcal{I}_{X,Y} = \mathcal{I}_X$ *s.j.*
5. $\mathcal{I}_{X,Y} \leq \mathcal{I}_X$ *s.j.*
6. $H(Y|X) = 0$
7. $H(Y|X) \leq 0$
8. $H(X,Y) = H(X)$
9. $H(X,Y) \leq H(X)$

*Důkaz.* Vzhledem k nezápornosti uvedených veličin jsou ekvivalentní podmínky (2) a (3), díky součtovým vzorcům pak (2) a (4) a také (3) a (5). Dostáváme tedy, že podmínky (2-5) jsou vzájemně ekvivalentní. Analogicky dostáváme vzájemnou ekvivalenci podmínek (6-9). Přechodem k střední hodnotě dostaneme z podmínek (2-5) podmínky (6-9). Navíc pokud střední hodnota nezáporné veličiny je nula, musí být i samotná veličina nulová skoro jistě. Tedy (6) implikuje (2). Získáváme tedy vzájemnou ekvivalenci podmínek (2-9). Nakonec dokážeme, že jsou ekvivalentní (1) a (5). Pokud $X \vdash Y$, pak je $X \vdash X, Y$. Z předchozího lemmatu pak dostáváme platnost podmínky (5). Naopak předpokládejme, že neplatí $X \vdash Y$. Potom $s(P_{X,Y})$ není zobrazením, t.j. existuje $(a,b), (a,b') \in s(P_{X,Y})$, $b \neq b'$. Pro $\omega \in (X,Y)^{-1}(a,b)$ platí

$$\mathbb{P}(X = a) \geq \mathbb{P}(X = a, Y = b) + \mathbb{P}(X = a, Y = b) > \mathbb{P}(X = a, Y = b).$$

Tedy $\mathcal{I}_{X,Y}(\omega)$ se liší od $\mathcal{I}_X(\omega)$. Jelikož $(X,Y)^{-1}(a,b)$ má kladnou pravděpodobnost, dostáváme, že pak neplatí (5). □

## 3.2 Nezávislost

Náhodné veličiny $X_1, X_2, ..., X_n$ jsou nezávislé právě když

$$P_{X_1,X_2,..,X_n}(a_1, a_2, ..., a_n) = P_{X_1}(a_1) P_{X_2}(a_2) \cdots P_{X_n}(a_n).$$

**Lemma 4.** *$X_1, X_2, .., X_n$ jsou nezávislé náhodné veličiny, (vektor budeme značit $X_{1..n}$), pak*

$$\mathcal{I}_{X_{1..n}} = \sum_{i=1}^{n} \mathcal{I}_{X_i} \text{ s.j.}, \qquad H(X_{1..n}) = \sum_{i=1}^{n} H(X_i).$$

*Důkaz.* Pokud jsou veličiny nezávislé, pak pro $\omega \in s(X_1, X_2, \ldots, X_n)$:

$$\begin{aligned} \mathcal{I}_{X_{1..n}}(\omega) &= -\log P_{X_{1..n}}(X_{1..n}(\omega)) = -\log \prod_{i=1}^{n} (P_{X_i}(X_i(\omega))) \\ &= \sum_{i=1}^{n} -\log(P_{X_i}(X_i(\omega))) = \sum_{i=1}^{n} \mathcal{I}_{X_i}(\omega). \end{aligned}$$

□

Rovnost pro informační obsah je přirozeně ekvivalentní nezávislosti veličin. Sumární vzorec pro entropii souboru veličin je v tuto chvíli důsledkem jejich nezávislosti (střední hodnoty veličin nemusí znamenat rovnost veličin s.j.). Později ale uvidíme, že sumární vzorec pro entropii také charakterizuje nezávislost.

**Příklad 1.** *Nechť* $X_1, X_2 : \Omega \to B$ *jsou nezávislé náhodné veličiny s rozdělením* $(\frac{1}{2}, \frac{1}{2})$ *a* $X_3 = (X_1 + X_2) \mod 2$.

Pak $X_i, X_j$ jsou nezávislé, kdykoliv $i \neq j$, ale trojice $X_1, X_2, X_3$ nezávislá není. Konkrétně každá dvojice determinuje třetí veličinu. Pro entropii platí (i,j různá)

$$H(X_i) = 1, \qquad H(X_i, X_j) = 2, \qquad H(X_1, X_2, X_3) = 2.$$

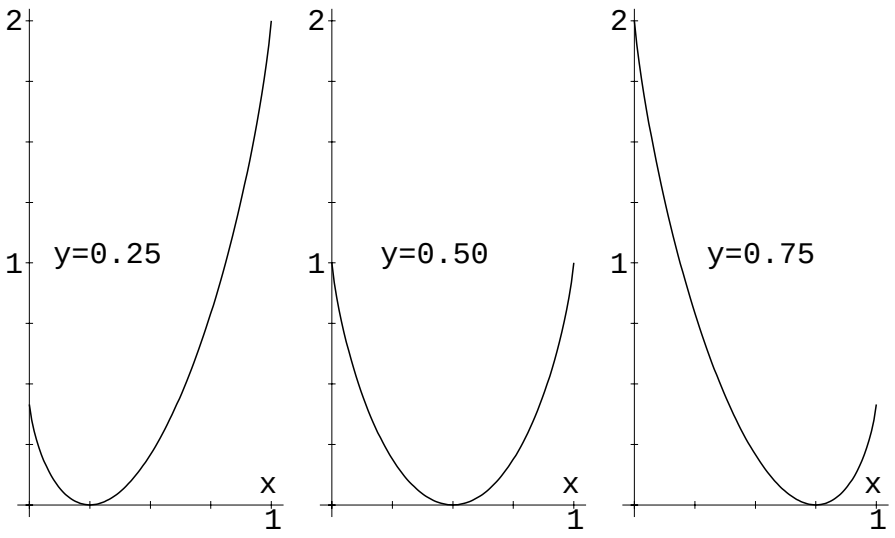
## 3.3 Divergence entropie

Uvažujme nyní dvě pravděpodobnostní rozdělení $P$ a $Q$ na stejné konečné množině $A$.

**Definice 2.** ***Divergence entropie*** *rozdělení* $P$ *vzhledem k rozdělení* $Q$ *je definována vzorcem*

$$D(P \parallel Q) = \sum_{a \in s(P)} P(a) \cdot \log \frac{P(a)}{Q(a)},$$

*pokud* $s(P) \subset s(Q)$. *Jinak,* $D(P \parallel Q) = +\infty$.




Obrázek 4: Divergence entropie

Na obrázku 4 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 divergence odvodíme z konvexity logaritmu.

Pro konvexní funkce na intervalu platí Jensenova nerovnost, kterou citujeme bez důkazu.

**Tvrzení 3** (Jensenova nerovnost I)**.** *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*

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

*Pokud je $f$ striktně konvexní a nastane rovnost, potom je množina $\{x_i \mid t_i > 0\}$ jednoprvková.*

**Tvrzení 4** (Jensenova nerovnost II)**.** *Buď $f : I \to \mathbb{R}$ konvexní na intervalu $I$, $\xi$ buď reálná náhodná veličina, pro kterou $\mathbb{P}(\xi \in I) = 1$, . Potom*

$$f(\mathbb{E}(\xi)) \leq \mathbb{E}(f(\xi)).$$

*Pokud je $f$ striktně konvexní a nastane rovnost, potom je $\xi$ triviální, t.j. $s(P_\xi)$ je jednoprvková.*

Ze striktní konkávnosti logaritmu plyne klíčová vlastnost divergence, totiž její nezápornost.

**Tvrzení 5.** *$D(P \parallel Q) \geq 0$ a rovnost nastává právě když $P = Q$.*

*Důkaz.* Stačí uvažovat případ $s(P) \subset s(Q)$. Jelikož je $-\log x$ striktně konvexní a striktně klesající, dostáváme

$$\begin{aligned} D(P \parallel Q) &= \sum_{a\in s(P)} P(a)\left(-\log\frac{Q(a)}{P(a)}\right) \geq -\log\left(\sum_{a\in s(P)} P(a)\cdot\frac{Q(a)}{P(a)}\right) \\ &= -\log\left(\sum_{a\in s(P)} Q(a)\right) \geq -\log\left(\sum_{a\in s(Q)} Q(a)\right) = 0. \end{aligned}$$

Pokud bychom chtěli rovnosti, namísto nerovností, musí být $s(P) = s(Q)$ a podíl $\frac{Q(a)}{P(a)}$ je roven konstantě $\alpha$, pro všechna $a \in s(P)$. Z předchozího plyne,

$$\alpha = \sum_{a\in s(P)} P(a)\frac{Q(a)}{P(a)} = \sum_{a\in s(Q)} Q(a) = 1.$$

□

Z předchozího tvrzení plyne řada užitečných výsledků.

**Tvrzení 6.** *Nechť má veličina $X$ hodnoty v $A$. Pak $H(X) \leq \log(\#A)$. Rovnost nastane právě tehdy když je $P_X$ rovnoměrně rozloženo na $A$.*

*Důkaz.* Na množině $A$ uvažujme rovnoměrné rozložení $U$, dané předpisem $U(a) = \frac{1}{\#A}$. Potom $s(P_X) \subset s(U)$ a

$$0 \leq D(P_X \parallel U) = \sum_{a\in s(P_X)} P_X(a)\log\left(\frac{P_X(a)}{\frac{1}{\#A}}\right) = \log(\#A) - H(X).$$

Kýžená rovnost nastane právě tehdy když $P_X$ a $U$ splývají. □

Následující věta dává divergenci do úzké souvislosti se vzájemnou informací. Jedná se tak o dalšího kandidáta na alternativní definici vzájemné entropie jako „divergence od nezávislosti“.

**Tvrzení 7.** *$H(X:Y) = D(P_{X,Y} \parallel P_X \cdot P_Y)$. Tedy $H(X:Y) \geq 0$ a rovnost nastává právě když $X$ a $Y$ jsou nezávislé.*

*Důkaz.* Pro náhodné veličiny $X$ a $Y$ s hodnotami v $A$ a $B$ uvažujme dvě rozdělení pravděpodobnosti na $A \times B$, konkrétně $P_{X,Y}$ a $(P_X \cdot P_Y)$, definované předpisem $(P_X \cdot P_Y)(a,b) = P_X(a)P_Y(b)$. Lze lehko nahlédnout, že $s(P_{X,Y}) \subset s(P_X \cdot P_Y)$. Proto

$$H(X:Y) = \sum_{(a,b)\in s(P_{X,Y})} P_{X,Y}(a,b) \log \frac{P_{X,Y}(a,b)}{P_X(a) \cdot P_Y(b)} = D(P_{X,Y} \parallel P_X \cdot P_Y).$$

Tedy $H(X:Y) \geq 0$ a rovnost nastane právě tehdy když $P_{X,Y}$ je totožné s $P_X \cdot P_Y$, neboli když jsou veličiny $X$ a $Y$ nezávislé. □

**Důsledek 1.** *Nechť $X, Y$ jsou náhodné veličiny. Pak*

*(1) $H(X,Y) \leq H(X) + H(Y)$ (subaditivita)*

*(2) $H(X|Y) \leq H(X)$. (monotónnost v podmínce, speciální případ)*

*Rovnosti nastanou právě tehdy, když jsou veličiny nezávislé.*

**Lemma 5.** *Nechť $X, Y, Z$ jsou náhodné veličiny. Pak*

*(1) $H(X,Y|Z) \leq H(X|Z) + H(Y|Z)$ (podmíněná subaditivita)*

*(2) $H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z)$ (submodularita)*

*(3) $H(X|Y,Z) \leq H(X|Z)$. (monotónnost v podmínce)*

*Důkaz.* Jednotlivé nerovnosti se dají na sebe převést pomocí součtových vzorců. Budeme dokazovat podmínku (2): Označme $H(X:Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z)$. Potom

$$\begin{aligned}
H(X:Y|Z) &= \sum_{(a,b,c)\in s(P_{X,Y,Z})} P_{X,Y,Z}(a,b,c) \left( -\log \frac{P_{X,Z}(a,c)P_{Y,Z}(b,c)}{P_{X,Y,Z}(a,b,c)P_Z(c)} \right) \\
&\geq -\log \sum_{(a,b,c)\in s(P_{X,Y,Z})} P_{X,Y,Z}(a,b,c) \frac{P_{X,Z}(a,c)P_{Y,Z}(b,c)}{P_{X,Y,Z}(a,b,c)P_Z(c)} \\
&= -\log \sum_{(a,b,c)\in s(P_{X,Y,Z})} \frac{P_{X,Z}(a,c)P_{Y,Z}(b,c)}{P_Z(c)} \\
&= -\log \sum_{(b,c)\in s(P_{Y,Z})} \frac{P_{Y,Z}(b,c)}{P_Z(c)} \sum_{a\in A:(a,b,c)\in s(P_{X,Y,Z})} P_{X,Z}(a,c) \\
&\geq -\log \sum_{(b,c)\in s(P_{Y,Z})} \frac{P_{Y,Z}(b,c)}{P_Z(c)} \cdot P_Z(c) \geq -\log 1 = 0.
\end{aligned}$$

Předposlední nerovnost plyne z toho, že sčítáme pravděpodobnosti disjunktních podmnožin množiny $Z^{-1}(c)$. □

Jiný způsob, jak vyjádřit monotónnost v podmínce je uvedena v následujícím lemmatu. Zajímavý je pak výsledek pro situaci, kdy $X$ a $Y$ determinují společnou (sdílenou) veličinu $Z$. Zde se potvrzuje očekávání, že je tato sdílená veličina "podmnožinou" vzájemné informace a tudíž je vzájemná informace $X$ a $Y$ větší rovna entropii "sdíleného klíče $Z$". Tato představa je ale jen určitou rozumnou intuicí. Ve-skutečnosti jde pouze o relaci mezi dvěma číselnými hodnotami. Vzájemná informace nemá žádnou jasnou strukturu, jen velikost.

**Lemma 6.** *Pro diskrétní náhodné veličiny $X, Y, Z$, $Y \vdash Z$ platí*

- $H(X|Y) \leq H(X|Z)$
- $H(X:Z) \leq H(X:Y)$
- $H(Z) \leq H(X:Y)$*, pokud navíc* $X \vdash Z$*.*

*Důkaz.*

$$H(X|Y) = H(X,Y) - H(Y) = H(X,Y,Z) - H(Y,Z) = H(X|YZ) \leq H(X|Z).$$

Druhá nerovnost plyne pomocí součtových vzorců z první. U třetí postupujeme následovně:

$$H(X:Y) \geq H(X:Z) \geq H(Z:Z) = H(Z).$$

□

# 4 Náhodné procesy

## 4.1 Základní definice

Podobně jako v klasické teorii pravděpodobnosti, teoretické pojmy začnou být "viditelné"na "datech" v momentě, kdy máme dat dostatek. To je princip statistiky, který je založen na četnostech výskytu určitých jevů, které pak dávají nahlédnout jejich pravděpodobnosti.

Celá teorie tedy nabývá zajímavějších obzorů až tehdy, zkoumáme-li namísto jedné veličiny, soubor veličin, či posloupnost veličin.

Uvažujme tedy posloupnost náhodných veličin $\mathbb{X} = X_0, X_1, X_2, \ldots$ definované na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ s hodnotami v konečné množině $A$, které budeme říkat abeceda procesu (v některých speciálních případech připustíme spočetnou abecedu). Konečné posloupnosti $u = u_0u_1u_2\ldots u_{n-1} \in A^n$ budeme nazývat slovy nad abecedou $A$. Délkou slova rozumíme délku dané posloupnosti, např. slovo $u = u_0u_1\ldots u_{n-1}$ má délku $n$. Značíme ji $|u|$. Množinu všech slov značíme $A^*$. Do této množiny také zahrneme prázdné slovo $\lambda$ délky 0.

Pro jednoduchost budeme také u posloupností všelikého druhu používat notaci:

$$\begin{aligned} X_k^{n-1} &= X_{[k,n)} = (X_k, X_{k+1}, \ldots, X_{n-1}), \\ a_k^{n-1} &= a_{[k,n)} = (a_k, a_{k+1}, \ldots, a_{n-1}), \\ a_k^{n-1} &= u_{[k,n)} = (u_k, u_{k+1}, \ldots, u_{n-1}). \end{aligned}$$

V první řadě nás bude zajímat jak roste entropie s přibývajícími veličinami, čili jak se chová $H(X_0^{n-1})$, $\mathcal{I}_{X_0^{n-1}}$, a jak tyto veličiny souvisí např. s kompresí. Definujeme **entropii procesu** předpisem

$$H(\mathbb{X}) = \lim_{n\to\infty} \frac{H(X_0^{n-1})}{n}.$$

Dále zaveďme

$$\mathcal{I}_{\mathbb{X}}(\omega) = \lim_{n\to\infty} \frac{\mathcal{I}_{X_0^{n-1}}(\omega)}{n}.$$

Z hlediska bezztrátové komprese se zajímáme o následující problém. Uvažujeme injektivní zobrazení $f : A^* \to \{0,1\}^*$. Kompresním poměrem, pro danou zprávu $u \in A^*$, pak myslíme $\frac{|f(u)|}{|u|}$.

Ukážeme, že entropie procesu je dobře definovaná pro třídu asymptoticky stacionárních procesů, do kterých patří např. i.i.d. procesy, či Markovské homogenní procesy, blokové procesy a další odvozené procesy. Pro ně lze potom říci, že nejlepší průměrný kompresní poměr, který můžeme dosáhnout při bezztrátové kompresi je roven entropii procesu, neboli že existuje injektivní $f : A^* \to \{0,1\}*$, takové, že

$$\limsup_{n\to\infty} \mathbb{E}\left(\frac{|f(X_0^{n-1})|)}{n}\right) = H(\mathbb{X}).$$

Zároveň ukážeme, že lepší komprese neexistuje.

Pro podtřídu procesů, takzvaně asymptoticky ergodických, dokážeme, že $\mathcal{I}_{\mathbb{X}}(\omega)$ je dobře definovaná pro skoro všechna $\omega$, a že se rovná konstantě $H(\mathbb{X})$. To nám umožní konstruovat kompresi, pro kterou bude kompresní poměr skoro jistě konvergovat k $H(\mathbb{X})$, nikoli jen v průměru, t.j.

$$\mathbb{P}\left(\omega : \lim_{n\to\infty} \frac{|f(X_0^{n-1}(\omega))|}{n} = H(\mathbb{X})\right) = 1.$$

Dokonce sestrojíme zobrazení, které bude mít tuto optimální kompresní vlastnost vůči jakémukoliv asymptoticky ergodickému procesu.

Typy procesů, které jsme definovali, jsou definovány následovně. Proces je

- i.i.d., pokud pro všechna $n \in \mathbb{N}$, $u \in A^n$,

$$\mathbb{P}(X_0^{n-1} = u) = \prod_{i=0}^{n-1} \mathbb{P}(X_0 = u_i),$$

- **stacionární**, pokud pro všechna $n, k \in \mathbb{N}$, $u \in A^n$,

$$\mathbb{P}(X_0^{n-1} = u) = \mathbb{P}(X_k^{n+k-1} = u),$$

- **ergodický**, pokud pro všechna $u \in A^n$, $v \in A^k$, platí

$$\mathbb{P}(X_0^{n-1} = u) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u | X_0^{k-1} = v),$$

  kdykoliv je $\mathbb{P}(X_0^{k-1} = v) > 0$.

- **asymptoticky stacionární**, pokud pro všechna $n \in \mathbb{N}$, $u \in A^n$, existuje limita průměrů

$$\frac{1}{m} \sum_{k=0}^{m-1} \mathbb{P}(X_k^{n+k-1} = u), \qquad m \to \infty,$$

- **asymptoticky ergodický**, pokud pro všechna $n$, $u \in A^n$, $v \in A^k$, existuje limita průměrů

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u | X_0^{k-1} = v),$$

  kdykoliv je

$$\limsup_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_0^{k-1} = v) > 0.$$

  Zároveň požadujeme, aby tato limita závisela pouze na $u$, nikoliv na $v$.

prověřit tuto podmínku

Pokud v definici asymptoticky ergodického procesu dosadíme za $v$ prázdné slovo, zjistíme, že asymptoticky ergodické procesy musí být asymptoticky stacionární. Stacionární proces je jistě asymptoticky stacionární a ergodický je jistě asymptoticky ergodický. Mimo jiné, limes superior z definice je pak roven klasické limitě. Ergodický systém je navíc stacionární. To lze nahlédnout, pokud zvolíme $v = \lambda$:

$$\begin{aligned}
\mathbb{P}(X_1^n = u) &= \sum_{a\in A} \mathbb{P}(X_0^n = au) = \sum_{a\in A} \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i} = au) \\
&= \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_{i+1}^{n+i} = u) \\
&= \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u) \\
&\quad + \lim_{m\to\infty} \frac{1}{m} \left(\mathbb{P}(X_m^{n+m-1} = u) - \mathbb{P}(X_0^{n-1} = u)\right) \\
&= \mathbb{P}(X_0^{n-1} = u) + 0
\end{aligned}$$

Nejsložitější je zřejmě nahlédnout následující vztah.

**Tvrzení 8.** *Pokud je asymptoticky ergodický proces stacionární, je ergodický.*

*Důkaz.* Mějme $n, k \in \mathbb{N}$, $u \in A^n$ a uvažujme všechny $v \in A^k$, pro které $P(X_0^{k-1} = v) > 0$. Díky stacionaritě je triviálně splněna podmínka na $v$ z definice asymptotické stacionarity. Existuje tedy nějaká konstanta $\alpha$, společná pro všechny $v \in A^k$, $\mathbb{P}(X_0^{k-1} = v) > 0$, tak že platí

$$\alpha = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u | X_0^{k-1} = v),$$

neboli

$$\alpha \cdot \mathbb{P}(X_0^{k-1} = v) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u, X_0^{k-1} = v).$$

Pokud nyní vysčítáme tyto rovnice přes všechny $v \in A^k$, pro které je $\mathbb{P}(X_0^{k-1} = v) > 0$, dostaneme

$$\alpha = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{P}(X_i^{n+i-1} = u) = \mathbb{P}(X_0^{n-1} = u).$$

V poslední rovnici jsme použili stacionaritu, tj. fakt, že všechny členy v sumě jsou stejné. Proces je tedy ergodický. □

Zatímco i.i.d. proces splňuje všechny vypsané podmínky, Markovský řetězec nemusí být ani stacionární, ani ergodický. Vždy je ovšem asymptoticky stacionární. Stacionarita řetězce ve smyslu naší definice je ekvivalentní se stacionaritou počátečního rozdělení. Asymptotická ergodicita je pak ekvivalentní s nerozložitelností množiny rekurentních stavů řetězce. Zároveň se dá ukázat, že se podobně budou chovat také funkce Markovských řetězců, alias skryté Markovské řetězce.

Na závěr kapitoly dokážeme, že mají entropii všechny stacionární procesy. Stejné tvrzení pro asymptoticky stacionární necháme na později. Nejprve dvě lemmata:

**Lemma 7.** *Pro nerostoucí posloupnost nezáporných reálných čísel $a_n$, $n \in \mathbb{N}$, existuje limita postupných průměrů $\frac{1}{n}\sum_{i=0}^{n-1} a_i$ a je rovna limitě vlastní posloupnosti.*

**Důkaz:** Označme $a = \lim_{n\to\infty} a_n$. Tato limita existuje, neboť posloupnost je monotónní a omezená. Zároveň platí $a \leq a_n$, $n \in \mathbb{N}$. Zvolme neklesající posloupnost přirozených čísel $k_n$, která jde do nekonečna, ale pomaleji než $n$, t.j. $\lim \frac{k_n}{n} = 0$ (například $k_n = \lfloor\sqrt{n}\rfloor$). Označme $c_n = \frac{1}{n}\sum_{i=0}^{n-1} a_i$. Potom platí

$$\begin{aligned} 0 \leq (c_n - a) &= \frac{1}{n}\sum_{i=0}^{k_n-1}(a_i - a) + \frac{1}{n}\sum_{i=k_n}^{n-1}(a_i - a) \\ &\leq \frac{1}{n}\sum_{i=0}^{k_n-1}(a_0 - a) + \frac{1}{n}\sum_{i=k_n}^{n-1}(a_{k_n} - a) \\ &\leq \frac{k_n(a_0 - a)}{n} + (a_{k_n} - a). \end{aligned}$$

Oba členy v posledním součtu jdou k nule, proto i $c_n$ konverguje k $a$. □

**Tvrzení 9.** *Nechť $X$ je stacionární proces.*

1. *Pro všechna $k, m, \ell \in \mathbb{N}$, platí*

$$s\left(P_{X_{[k,m)}}\right) = s\left(P_{X_{[k+\ell,m+\ell)}}\right), \qquad H(X_{[k+\ell,m+\ell)}) = H(X_{[k,m)}),$$

2. *posloupnost $H(X_n|X_{[0,n)})$, $n \in \mathbb{N}$, je nerostoucí,*

3. *entropie procesu $H(X)$ je dobře definovaná a je rovna limitě*

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

**Důkaz:** Buď $X$ stacionární proces $k, m, \ell \in \mathbb{N}$, $k < m$, $u \in A^{m-k}$. Z definice stacionarity dostáváme, že se rovnají pravděpodobnosti $\mathbb{P}(X_{[k,m)} = u)$ a $\mathbb{P}(X_{[k+\ell,m+\ell)} = u)$. Z toho plyne $s(P_{X_{[k,m)}}) = s(P_{X_{[k+1,m+1)}})$ a

$$\begin{aligned} H(X_{[k,m)}) &= \sum_{u\in s(P_{X_{[k,m)}})} -\mathbb{P}(X_m^{k-1} = u)\log\mathbb{P}(X_m^{k-1} = u) \\ &= \sum_{u\in s(P_{X_{[k+\ell,m+\ell)}})} -\mathbb{P}(X_{m+\ell}^{k-1+\ell} = u)\log\mathbb{P}(X_{m+\ell}^{k-1+\ell} = u) \\ &= H(X_{[k+\ell,m+\ell)}). \end{aligned}$$

Položme $a_n := H(X_n|X_{[0,n)})$. Z právě dokázaného tvrzení a z Lemmatu 5 plyne

$$a_n \leq H(X_n|X_{[1,n)}) = H(X_{[1,n+1)}) - H(X_n) = H(X_{[0,n)}) - H(X_{n-1}) = a_{n-1}.$$

Posloupnost $a_n$ je tedy nerostoucí. Jistě je i nezáporná. Z řetězového pravidla dostáváme, že

$$\frac{1}{n}H(X_{[0,n)}) = \frac{1}{n}\sum_{i=0}^{n-1} a_i.$$

K dokončení důkazu podmínky (3) pak stačí aplikovat předchozí lemma. □

## 4.2 Rozdělení náhodného procesu

Podobně, jako v jiných partiích teorie pravděpodobnosti, i v případě teorie informace hraje hlavní roli rozdělení náhodných veličin, potažmo rozdělení náhodného procesu. Uvidíme, že naprostá většina našich tvrzení a vět nezmění svou platnost, když nahradíme uvažovaný proces, jiným procesem se stejným rozdělením.

Připomeňme, že rozdělení náhodné veličiny $X : \Omega \to \mathbb{R}$, je pravděpodobnostní míra $P_{\mathbb{X}}$ na Borelovských množinách $\mathcal{B}(\mathbb{R})$, která je definovaná předpisem $P_X(U) = \mathbb{P}(X^{-1}(U))$, pro všechna $U \in \mathcal{B}(\mathbb{R})$. Podobně, pro obecný "náhodný objekt", neboli pro měřitelné zobrazení $X$ z $(\Omega, \mathcal{F})$ do $(\Omega', \mathcal{F}')$ je rozložením $X$ pravděpodobnostní míra $P_X$, kde $P_X(U) = \mathbb{P}(X^{-1}U)$, pro všechna $U \in \mathcal{F}'$. Do této velmi obecné definice se vejde například jednoduchý případ námi uvažovaných diskrétních veličin $X : \Omega \to A$, kde $A$ je konečná množina. Měřitelnými podmnožinami $A$ jsou pak všechny její podmnožiny a míra je plně charakterizovaná svými hodnotami na jednoprvkových množinách $P_X(\{a\})$, $a \in A$, kterou my značíme pro jednoduchost $P_X(a)$.

My můžeme tuto definici ale také uplatnit na celý náhodný proces, pokud ho chápeme, jako zobrazení $\mathbb{X} : \Omega \to A^{\mathbb{N}}$. Z měřitelnosti jednotlivých $X_i$, $i \in \mathbb{N}$, plyne měřitelnost $\mathbb{X}$ vůči $\mathcal{B}(A^{\mathbb{N}})$, což je nejmenší $\sigma$-algebra obsahující množiny

$$\{x \in A^{\mathbb{N}} \mid x_i = a\}, \qquad i \in \mathbb{N}, a \in A.$$

Z uzavřenosti na průniky plyne, že jsou měřitelné také všechny množiny

$$[u] := \{x \in A^{\mathbb{N}} \mid x_0^{|u|-1} = u\}, \qquad u \in A^*.$$

Tyto množiny nazýváme cylindry. Systém cylindrů tvoří množinovou algebru, t.j. systém je uzavřený na konečné průniky a každá množina se dá doplnit na nejvýše spočetný systém vzájemně disjunktních množin, jejichž sjednocením je $A^{\mathbb{N}}$. Díky tomu je rozdělení procesu $P_X$ jednoznačně určeno hodnotami

$$P_{\mathbb{X}}([u]) = \mathbb{P}(\mathbb{X}^{-1}([u])) = \mathbb{P}(X_0^{n-1} = u_0^{n-1}),$$

$n \in \mathbb{N}$, $u \in A^n$.

Kromě rozdělení $P_{\mathbb{X}}$ na $A^{\mathbb{N}}$, zaveďme také zobrazení $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$, které má předpis

$$(\sigma((x_i)_{i\in\mathbb{N}}))_j = x_{j+1}, \qquad x \in A^{\mathbb{N}}, j \in \mathbb{N}.$$

Zobrazení posouvá souřadnice, proto ho nazýváme "posunem". Pomocí tohoto zobrazení a příslušného rozdělení můžeme vyjádřit vlastnosti procesu z této kapitoly i potřebná tvrzení o entropii. V dalších kapitolách se budeme zabývat právě strukturou $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), P_{\mathbb{X}}, \sigma)$.

**Příklad 2.** *Vyjádřete vlastnosti procesů z této kapitoly pomocí rozdělení $P_{\mathbb{X}}$ a zobrazení $\sigma$.*

**Příklad 3.** *Dokažte, že pokud mají dvě náhodné veličiny $X$ a $Y$ stejné rozdělení, mají stejné rozdělení také informační obsahy $\mathcal{I}_X$ a $\mathcal{I}_Y$. Zároveň mají stejnou entropii.*

**Příklad 4.** *Dokažte, že pokud mají dva procesy $\mathbb{X}$ a $\mathbb{Y}$ stejné rozdělení, mají i stejné vlastnosti a stejnou entropii.*

**Příklad 5.** *Zkonstruujte příklad, kdy mají veličiny $\mathcal{I}_X$ a $\mathcal{I}_Y$ stejné rozdělení, ale $\mathbb{P}(\mathcal{I}_X = \mathcal{I}_Y) = 0$.*

# 5 Dynamické systémy, definice a příklady

Pro vhodný formální popis $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), P_{\mathbb{X}}, \sigma)$ z předchozí kapitoly, zavedeme pojem dynamického systému a uvedeme některé vybrané části ergodické teorie.

Dynamické systémy mají modelovat vývoj složitých systémů v čase. V případě tohoto textu budeme uvažovat diskrétní časové okamžiky, tedy čas bude reprezentován diskrétní množinou a to převážně množinou přirozených čísel $\mathbb{N}$, případně množinou celých čísel $\mathbb{Z}$. Druhý příklad umožňuje záporný čas, který interpretujeme jako minulost systému. Poznamenejme, že do množiny $\mathbb{N}$ řadíme nulu.

Budeme tedy uvažovat množinu $\Omega$, zobrazení $T : \Omega \to \Omega$ a jeho iterace $T^n$, definované induktivně předpisem

$$T^0 = Id_\Omega, \qquad T^{n+1} = T \circ T^n, \qquad T^{-n} = (T^{-1})^n,$$

kde $n$ je přirozené a $T^{-n}$ je definováno pouze v případě, že je $T$ bijekce na $\Omega$ (toto většinou předpokládat nebudeme). Máme-li bod $x \in \Omega$, potom $T^n(x)$ budeme považovat za stav systému v čase $n$, za předpokladu, že systém "začínal" v bodě $x$. Posloupnost $(T^n(x))_{n\in\mathbb{N}}$ budeme nazývat trajektorií bodu $x$, množinu bodů $T^n(x)$, $n \in \mathbb{N}$, budeme nazývat orbitou tohoto bodu. Pokud bude existovat $T^{-1}$ zavedeme analogicky pojem oboustranné trajektorie $(T^n(x))_{n\in\mathbb{Z}}$ a oboustranné orbity. Pokud platí $x = T(x)$, řekneme, že $x$ je pevným (fixním) bodem, pokud platí $x = T^n(x)$, nazveme bod periodickým a $n$ nazýváme periodou bodu. Většinou pak mluvíme o nejmenším takovém $n$ pro daný bod. Množinu $B \subset \Omega$ nazveme invariantní, pokud $B = T^{-1}B$, slabě invariantní, pokud $B \subset T^{-1}(B)$. Na množině $\Omega$ budeme uvažovat $\sigma$-algebru měřitelných množin a pravděpodobnostní míru. Zajímat nás bude evoluce takové míry.

**Definice 3.** *Měřitelných dynamickým systémem nazveme trojici $(\Omega, \mathcal{B}, T)$, kde $\mathcal{B}$ je $\sigma$-algebra podmnožin $\Omega$ a $T$ je měřitelné zobrazení z $\Omega$ do $\Omega$. Pravděpodobnostním dynamickým systémem nazveme čtveřici $(\Omega, \mathcal{B}, \mu, T)$, kde $(\Omega, \mathcal{B}, T)$ je měřitelný dynamický systém a $\mu$ je pravděpodobnostní míra na $\mathcal{B}$. Řekneme, že je tento systém invariantní, pokud je $\mu$ $T$-invariantní , t.j. pokud $\mu(T^{-1}B) = \mu(B)$ pro každé $B \in \mathcal{B}$.*

*Pokud navíc $T$ zobrazuje $\Omega$ bijektivně a oboustranně měřitelně na $\Omega$, říkáme že je dynamický systém invertibilní.*

V případě pravděpodobnostního dynamického systému budeme obvykle předpokládat, že $\mu$ je úplná. Pokud tomu tak nebude z definice, zúplníme $\sigma$-algebru $\mathcal{B}$. Pokud uvažujeme různé míry na stejném měřitelném prostoru, musíme brát v potaz, že se v takovém případě po zúplnění i podkladové měřitelné dynamické systémy začnou lišit. Ve velké části tohoto textu budeme pracovat primárně s invariantním pravděpodobnostním dynamickým systémem, t.j. s invariantní mírou.

Bude-li jasné, kterou $\sigma$-algebru máme na mysli, budeme zápis zkracovat na $(\Omega, \mu, T)$, podobně v případě míry $\mu$ a zobrazení $T$. Zároveň budeme v případě obou typů dynamických systémů vynechávat upřesnění, zda jde o měřitelný, či pravděpodobnostní dynamický systém, pokud to bude jasné ze zápisu a kontextu.

Na zobrazení $T$ budeme hledět jako na zobrazení pravděpodobnostního prostoru $(\Omega, \mathcal{B}, \mu)$ do sebe sama. Takové zobrazení se obvykle nazývá *endomorfismem*. Pokud je navíc zobrazení $T$ bijektivní a $T^{-1}$ je endomorfismem na $\Omega$, nazývá se $T$ *automorfismem* nebo *transformací* nebo invertibilním systémem. Zobrazení $T$ budeme uvažovat jakožto morfismus v kategorii pravděpodobnostních prostorů, tedy nebude žádným problémem, pokud bude definován pouze $\mu$-skoro všude, invertovatelný též jen $\mu$-skoro všude. Dvě zobrazení pro nás budou totožná, pokud budou shodná skoro všude.

Kromě měřitelného systému budeme uvažovat ještě systém topologický.

**Definice 4.** *Topologickým systémem nazveme trojici $(\Omega, \mathcal{G}, T)$, kde $\mathcal{G}$ je kompaktní Hausdorffova topologie taková, že $T$ je vzhledem k této topologii spojité zobrazení z $\Omega$ do $\Omega$ (tedy $T^{-1}G \in \mathcal{G}$ pro všechny $G \in \mathcal{G}$).*

Vazba mezi topologickým a měřitelným systémem je následující.

**Věta 1** (Krylov-Bogoljubov)**.** *Buď $(\Omega, \mathcal{G}, T)$ topologický dynamický systém. Pak existuje pravděpodobnostní míra $\mu$ na Borelovských množinách $\mathcal{B}(\mathcal{G})$, která je invariantní vůči $T$. Tedy $(\Omega, \mathcal{B}(\mathcal{G}), \mu, T)$ je invariantní pravděpodobnostní dynamický systém.*

Invariantních měr může být více. Mezi nimi nás budou zajímat ty, které splňují následující definici.

**Definice 5.** *Systém $(\Omega, \mathcal{B}, \mu, T)$ nazveme ergodickým, pokud každá měřitelná invariantní množina je triviální, t.j. každá množina $A \in \mathcal{B}$ splňující $A = T^{-1}A$ má míru 0 nebo 1.*

Dříve než uvedeme příklady vyslovme důležité lemma a jeho důsledek.

**Lemma 8.** *Nechť $(\Omega, \mathcal{B}, \mu)$ je pravděpodobnostní prostor a $T$ měřitelné zobrazení na něm. Potom množina*

$$\mathcal{B}' = \{B \in \mathcal{B} \mid \mu(B) = \mu(T^{-1}B)\}$$

*tvoří Dynkinův systém, t.j. neprázdný soubor podmnožin $\Omega$ uzavřený na doplňky a spočetná disjunktní sjednocení.*

*Důkaz.* Prázdná množina a celé $\Omega$ jistě náleží do $\mathcal{B}'$. Pokud $B \in \mathcal{B}'$, pak

$$\begin{aligned}\mu(T^{-1}(\Omega \setminus B)) &= \mu(T^{-1}\Omega \setminus T^{-1}B) = \mu(\Omega \setminus T^{-1}B) = 1 - \mu(T^{-1}B) \\ &= 1 - \mu(B) = \mu(\Omega \setminus B).\end{aligned}$$

Pokud $B_1, B_2, \ldots$ jsou po dvou disjunktní množiny z $\mathcal{B}'$, potom jejich vzory jsou také po dvou disjunktní a platí

$$\begin{aligned}\mu(T^{-1}(\bigcup_{i\in\mathbb{N}} B_i)) &= \mu(\bigcup_{i\in\mathbb{N}} T^{-1}(B_i)) = \sum_{i\in\mathbb{N}} \mu(T^{-1}(B_i)) \\ &= \sum_{i\in\mathbb{N}} \mu(B_i) = \mu(\bigcup_{i\in\mathbb{N}} B_i).\end{aligned}$$

□

**Důsledek 2.** *Nechť $(\Omega, \mathcal{B}, \mu)$ je pravděpodobnostní prostor a $T$ měřitelné zobrazení na něm. Nechť $\mathcal{A}$ je soubor měřitelných množin generující $\mathcal{B}$, který je uzavřený na konečné neprázdné průniky. Pokud platí podmínka $\mu(B) = \mu(T^{-1}B)$ pro všechny množiny z $\mathcal{A}$, pak tato podmínka platí pro všechny množiny z $\mathcal{B}$.*

Tento důsledek plyne bezprostředně z předchozího lemmatu a Dynkinovy věty.

**Příklad 1** (Rotace kružnice)**.** Buď $\mathbb{T} = \{x \in \mathbb{C} \mid |x| = 1\}$ jednotková kružnice v komplexní rovině, opatřená topologií indukovanou z komplexní roviny. Pro oblouk $L$, označme $\ell(L)$ délku oblouku L. Takto definovaná funkce $\ell$ se dá jednoznačné rozšířit na míru na Borelovských množinách $\mathcal{B}(\mathbb{T})$. Tato míra, stejně jako její normalizovaná verze $\mu$, je invariantní vůči otáčení $R_\phi(x) = x \cdot e^{i\phi}$, pro všechna $\phi \in \mathbb{R}$. Dostáváme, že $(\mathbb{T}, \mathcal{B}(\mathbb{T}), \mu, R_\phi)$ je invariantním dynamickým systémem, který je navíc invertibilní a ergodický. Otáčení je spojité, takže dostáváme též topologický systém.

**Příklad 2** (Zdvojení kružnice)**.** Zdvojení kružnice je systém definovaný stejně jako v předchozím příkladu, jen zobrazení je definované předpisem $T(x) = x^2$. Takovéto zobrazení je opět spojité, definuje tudíž topologický dynamický systém, zachovává míru definovanou v předešlém bodě a příslušný pravděpodobnostní systém je ergodický. Zobrazení ovšem není invertibilní (a to ani skoro všude).

**Příklad 3** (Stan)**.** Buď $\mathbb{I} = [0, 1]$ jednotkový interval a $T$ zobrazení definované na $\mathbb{I}$ předpisem

$$T(x) = 1 - |1 - 2x|.$$

Topologický systém $(\mathbb{I}, T)$ se nazývá "stan". Invariantních měr je tu více. Není těžké ukázat, ze klasická Lebesgueova na intervalu je invariantní a ergodická. Mimo to má zobrazení pevný bod, tj. bod pro který $x = T(x)$. Takovým bodem jsou $2/3$. Druhou invariantní mírou je tedy Diracova míra soustředěná v tomto bodě. Tato míra je také ergodická. Invariantní, nikoliv však ergodická, je též jakákoliv konvexní kombinace těchto měr.

**Příklad 4** (Zdvojení intervalu)**.** Buď $\mathbb{I} = [0, 1[$ jednotkový interval a $T$ zobrazení definované na $\mathbb{I}$ předpisem

$$T(x) = 2x \mod 1.$$

Invariantních měr je opět více. Není těžké ukázat, ze klasická Lebesgueova na intervalu je invariantní a ergodická. Mimo to má zobrazení pevný bod, tj. bod pro který $x = T(x)$. Takovým bodem jsou $2/3$. Druhou invariantní mírou je tedy Diracova míra soustředěná v tomto bodě. Tato míra je také ergodická. Invariantní, nikoliv však ergodická, je též jakákoliv konvexní kombinace těchto měr.

**Definice 6.** *Nechť $(\Omega, \mathcal{A}, \mu, T)$ a $(\Gamma, \mathcal{B}, \nu, S)$ jsou dva dynamické systémy. Homomorfismem jednoho systému do druhého je jakékoliv měřitelné zobrazení $\phi$ z množiny $\Omega' \in \mathcal{A}$ plné míry do $\Gamma$ takové, že*

$$(\phi \circ T)|_{\Omega'} = (S \circ \phi)|_{\Omega'}, \quad a \quad \mu(\phi^{-1}(A)) = \nu(A),$$

*pro každé $A \in \mathcal{B}$.*

*Zobrazení $\phi$ je isomorfismem, pokud zobrazí $\Omega'$ bijektivně, oboustranně měřitelně, na množinu $\Gamma' \in \mathcal{B}$ plné míry.*

Existuje-li mezi dvěma systémy $(\Omega, \mathcal{A}, \mu, T)$ a $(\Gamma, \mathcal{B}, \nu, S)$ isomorfismus, řekneme, ze jsou systémy isomorfní. Existuje-li homomorfismus z $\Omega$ do $\Gamma$, pak $(\Gamma, \mathcal{B}, \nu, S)$ je faktorem $(\Omega, \mathcal{A}, \mu, T)$ a $(\Omega, \mathcal{A}, \mu, T)$ je rozšířením $(\Gamma, \mathcal{B}, \nu, S)$. Zobrazení $\phi$ se v takovém případě nazývá též faktorizujícím zobrazením.

**Lemma 9.** *Nechť $(\Omega, \mathcal{A}, \mu, T_1)$ je dynamický systém a $T_2$ je měřitelné zobrazení definované na $\Omega_0$ plné míry shodné s $T_1$ skoro všude. Potom $(\Omega, \mathcal{A}, \mu, T_2)$ je také dynamický systém, který je isomorfní s původním. Isomorfismem mezi těmito systémy je identita definovaná na příslušné množině plné míry.*

**Cvičení 1.** *Zdvojení kružnice je isomorfní se zdvojením intervalu.*

**Cvičení 2.** *Rotace kružnice $R_\phi$je isomorfní se systémem $(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda, S)$, kde $S(x) = (x + \phi)$* mod 1*.*

**Cvičení 3.** *Rotace kružnice $R_{2\phi}$ je faktorem rotace kružnice $R_\phi$.*

# 6 Symbolické systémy

Uvažujme nějaký konečný diskrétní prostor $A$ a jeho Kartézskou mocninu $A^{\mathbb{Z}}$, resp. $A^{\mathbb{N}}$ opatřenou produktovou topologií. Symbolickým systémem nazveme každý topologický systém definovaný na tomto topologickém prostoru. Typickým symbolickým systémem je jednostranný a oboustranný posun. Jednostranný posun $\sigma : A^{\mathbb{N}} \longrightarrow A^{\mathbb{N}}$ je definován předpisem

$$\sigma(a_0 a_1 a_2 \ldots) = a_1 a_2 \ldots$$

a oboustranný $\sigma : A^{\mathbb{Z}} \longrightarrow A^{\mathbb{Z}}$ předpisem

$$\sigma(\ldots a_{-2} a_{-1} . a_0 a_1 a_2 \ldots) = \ldots a_{-1} a_0 . a_1 a_2 a_3 \ldots,$$

kde tečka v zapíšu oboustranné posloupnosti z $A^{\mathbb{Z}}$ bude vždy označovat místo, které dělí indexy na záporné a nezáporné. Taková zobrazení jsou spojitá a na. Oboustranný posun je invertibilní a jeho inverze je též spojitá. Navíc jsou oba metrizovatelné následujícími metrikami $d_{\mathbb{Z}}$ resp. $d_{\mathbb{N}}$:

$$\begin{aligned} d_{\mathbb{N}}(x,y) &= 2^{\min\{i \geq 0, x_i \neq y_i\}} \\ d_{\mathbb{Z}}(x,y) &= 2^{\min\{i \geq 0, x_i \neq y_i \text{ nebo } x_{-i} \neq y_{-i}\}} \end{aligned}$$

pro $x \neq y$ a $d_{\mathbb{N}}(x,x) = 0$, $d_{\mathbb{Z}}(x,x) = 0$.

**Definice 7.** *Čtveřici $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu, \sigma)$, kde $\mu$ je pravděpodobnostní Borelovská míra, nazveme procesem.*

Tento pojem je volen s ohledem na souvislost s klasickým náhodným procesem popsanou níže.

Struktura symbolických systémů se často popisuje pomocí cylindrů a slov. Slovem nazveme každou konečnou posloupnost prvků z $a$, indexovanou čísly $0, 1, \ldots, n-1$, kde $n$ je délka slova. Označme $A^n$ množinu slov délky $n$, $\lambda$ bud prázdné slovo, jediné slovo délky 0. Dále bud $A^* = \bigcup_{n \in \mathbb{N}} A^n$ množina všech slov. Pro slovo $u \in A^*$, $|u|$ značí jeho délku. Slova nad abecedou $A$ tvoří monoid vzhledem k operaci konkatenace. Konkatenaci dvou slov $u$ a $v$ značíme $uv$. Je to slovo délky $|u| + |v|$, definované předpisem $uv_i = u_i$, pro $i < |u|$ a $uv_i = v_{i-|u|}$, pro $|u| \leq i < |u| + |v|$. Jednotkovým prvkem monoidu je prázdné slovo. Množinu $A$ nazýváme abecedou. Pokud $w = uvz$, pak všechna slova $u, v$ i $z$ nazýváme podslovy slova $w$. Slovo $u$ nazýváme prefixem slova $w$ a slovo $w$ nazýváme pravým rozšířením slova $u$. Relaci podslova a prefixu značíme $u \sqsubseteq w$ resp. $u \preceq w$. Pokud $u \neq w$, pak říkáme, že je prodloužení, či podslovo vlastní.

V jednostranném symbolickém prostoru definujme pro slovo $u \in A^*$ cylindr

$$[u] = \{(a_i)_{i \in \mathbb{N}}, a_i = u_i, \text{ pro všechna } i = 0, 1, \ldots, |u| - 1\}$$

V oboustranném symbolickém prostoru definujeme pro slova $u, v \in A^*$ cylindr následovně

$$[v.w] = \{(x_i)_{i \in \mathbb{Z}}, x_{-|v|} x_{-|v|+1} \ldots x_{-1} = v, x_0 x_1 \ldots x_{n-1} = w\}$$

Platí $[u]_n = [v.w]$, kde $v = u_0 u_1 \ldots u_{|u|+n-1}$ a $w = u_{|u|+n} \ldots u_{|u|+1}$.

Cylindry tvoří baží obojetnou bázi $A^{\mathbb{N}}$ resp. $A^{\mathbb{Z}}$ i bázi $\sigma$-algebry Borelovských množin. Navíc platí,

- $[v] \subseteq [u]$ tehdy a jen tehdy je-li $u$ prefixem slova $v$,
- $[v]$ a $[u]$ mají neprázdný průnik tehdy a jen tehdy je-li jedno ze slov prefixem druhého.

Přirozeně platí, že $[v] \subseteq [u]$ tehdy a jen tehdy je-li $u$ prefixem slova $v$.

**Věta 2** (Kolmogorovova věta o rozšíření)**.** *Každá nezáporná konečně aditivní reálná funkce zadaná na cylindrech se dá jednoznačné rozšířit na konečnou míru na $A^{\mathbb{Z}}$ resp. $A^{\mathbb{N}}$.*

Následující lemma ukazuje, že konečnou aditivitu v případě $A^{\mathbb{N}}$ lze ověřit pomocí podmínky, která se nazývá podmínka konzistence.

**Lemma 10** (Podmínka konzistence)**.** *Reálná funkce na cylindrech je konečně aditivní tehdy a jen tehdy platí-li následující podmínka:*

$$\mu([u]) = \sum_{a \in A} \mu([ua]), \qquad \textit{pro všechna slova } u.$$

*Důkaz.* Jistě z konečné aditivity vyplývá podmínka konzistence. Z podmínky konzistence lze lehce dokázat indukcí podle $n$, že pro každé $n$,

$$\mu([u]) = \sum_{v \in A^n} \mu([uv]), \qquad u \in A^*.$$

Buď nyní $u \in A^*$, $M \subset A^*$ konečná množina, taková, že cylindr $[u]$ je disjunktním sjednocením cylindrů $[v]$, $v \in M$. Položme $q = \max\{|v|, v \in M\}$,

$$S_1 = \sum_{v \in M} \mu([v]) = \sum_{v \in M} \sum_{w \in A^{q-|v|}} \mu([vw]).$$

Slova $vw$ z předchozí sumy jsou po dvou různá, mají délku $q$ a jsou prodloužením slova $u$, protože $[v] \subseteq [u]$. Tato slova splývají se souborem všech možných prodloužení slova $u$ délky $q$. Je-li totiž $w'$ prodloužení slova $u$, tak $[w']$ je obsažen v $[u]$ a proto musí mít neprázdný průnik s některým cylindrem $[v]$, $v \in M$. Příslušné slovo $v$ musí být prefixem $w'$, nebo naopak. Jelikož $|w'| \geq |v|$, je $v$ prefixem $w'$, tedy $w'$ je tvaru $vw$. Proto

$$S_1 = \sum_{w' \in A^q, u \preceq w'} \mu([w']) = \sum_{v' \in A^{q-|u|}} \mu([uv']) = \mu([u]).$$

□

**Lemma 11** (Podmínka invariance)**.** *Míra na $A^{\mathbb{N}}$ je invariantní tehdy a jen tehdy platí-li následující podmínka:*

$$\mu([u]) = \sum_{a \in A} \mu([au]), \qquad \textit{pro všechna slova } u.$$

*Důkaz.* Podmínka v lemmatu je shodná s podmínkou, že míra vzoru cylindru je totožná s mírou cylindru. Cylindry jsou uzavřené na průniky a generují $\sigma$-algebru. Můžeme tedy aplikovat Důsledek 2. □

Vraťme se nyní k souvislosti s náhodnými procesy. Označme projekce z $A^{\mathbb{N}}$ do jednotlivých souřadnic jako $\pi_n$, tedy $\pi_n : A^{\mathbb{N}} \to A$ je definovaná předpisem $\pi_n(x) = x_n$, pro $x \in A^{\mathbb{N}}$. Na tyto zobrazení můžeme nahlížet jako na náhodné veličiny definované na stejném pravděpodobnostním prostoru $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu)$, tedy to je náhodný proces s hodnotami v $A$.

Připomeňme, že Zmíněný proces $(\pi_i)_{i \in \mathbb{N}}$ má zjevně rozdělení rovné původní míře $\mu$, neboť z definice projekce plyne

$$\mu(\pi_{[0,n)} = u) = \mu([u]), \qquad u \in A^*.$$

Tento fakt shrnuje a rozšiřuje následující věta.

**Tvrzení 10.** *Buď $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu)$ pravděpodobnostní prostor. Potom projekce $(\pi_i)_{i\in\mathbb{N}}$ tvoří náhodný proces s rozdělením $\mu$.*

*Je-li $\mathbb{X} = (X_i)_{i\in\mathbb{N}}$ náhodný proces na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ s hodnotami v konečné množině $A$, je $(\pi_i)_{i\in\mathbb{N}}$ definovaný na $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), P_{\mathbb{X}})$ proces se stejným rozdělením.*

Z definic a z faktu, že invarianci lze ověřovat pouze na cylindrech, plyne také následující tvrzení.

**Tvrzení 11.** *Náhodný proces $\mathbb{X} = (X_i)_{i\in\mathbb{N}}$ s hodnotami v konečné množině $A$, je stacionární, právě tehdy když je systém $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), P_{\mathbb{X}}, \sigma)$ invariantní.*

Postupně uvidíme, že také ergodicita procesu odpovídá ergodicitě systému a zároveň zavedeme odpovídající pojmy pro asymptotickou stacionaritu a asymptotickou ergodicitu.

Jelikož námi zkoumané vlastnosti procesů závisí jen na vlastnostech odvozených dynamických systémů, které pro jednoduchost nazýváme také procesy, budeme jednotlivé příklady procesů rovnou popisovat pomocí rozdělení.

**Příklad 5** (Bernoulliovský posun, i.i.d. proces)**.** Buď $(A^{\mathbb{Z}}, \sigma)$ oboustranný posun a buď $(p_a)_{a\in A}$ pravděpodobnostní vektor. Pro slova $u$ a $v$ délky $m$ a $n$ a příslušný cylindr definujme reálnou funkci $\mu$ následovně:

$$\mu([u.v]) = (p_{u_0} p_{u_1} \cdot \ldots \cdot p_{u_{m-1}})(p_{v_0} p_{v_1} \cdot \ldots \cdot p_{v_{n-1}}).$$

Tato funkce se dá jednoznačně rozšířit na pravděpodobnostní míru na Borelovských množinách na $A^{\mathbb{Z}}$, která je invariantní vůči posunu. Čtveřice $(A^{\mathbb{Z}}, \mathcal{B}(A^{\mathbb{Z}}), \mu, \sigma)$ je ergodickým dynamickým systémem. Stejná konstrukce funguje pro jednostranný posun.

**Příklad 6** (Markovský posun, Markovský proces)**.** Buď $(A^{\mathbb{Z}}, \sigma)$ oboustranný posun, $(M_{a,b})_{a,b\in A}$ stochastická matice, tj. matice s nezápornými koeficienty jejíž řádky se sečtou na jednotku. Z Perron-Frobeniovy věty plyne, že existuje pravděpodobnostní vektor $(p_a)_{a\in A}$ takový, že $pM = p$. Vezmeme takový vektor a pro slova $u$ a $v$ délky $m$ a $n$ a příslušný cylindr definujme reálnou funkci $\mu$ následovně:

$$\mu([u.v]) = p_{u_0}(M_{u_0,u_1} M_{u_1,u_2} \cdot \ldots \cdot M_{u_{m-2},u_{m-1}}) M_{u_{m-1},v_0} (M_{v_0,v_1} M_{v_1,v_2} \cdot \ldots \cdot M_{v_{n-2},v_{n-1}})$$

Tato funkce se dá jednoznačně rozšířit na pravděpodobnostní míru na Borelovských množinách na $A^{\mathbb{Z}}$, která je invariantní vůči posunu. Čtveřice $(A^{\mathbb{Z}}, \mathcal{B}(A^{\mathbb{Z}}), \mu, \sigma)$ je dynamickým systémem, ne vždy ergodickým.

Stejná konstrukce funguje pro jednostranný posun, kde navíc můžeme zvolit libovolný pravděpodobnostní vektor $p$ (nemusí splňovat $pM = p$).

Definujeme

$$\mu([u]) = p_{u_0}(M_{u_0,u_1} M_{u_1,u_2} \cdot \ldots \cdot M_{u_{n-2},u_{n-1}}).$$

**Cvičení 4.** *Dokažte podmínku konzistence a invariance pro Markovské a Bernoulliovské jednostranné procesy.*

# 7 Ergodické systémy

V dynamickém systému $(\Omega, \mathcal{A}, \mu, T)$ bývá důležité, jak se chovají trajektorie jednotlivých bodů, kde trajektorií bodu $x$ rozumíme posloupnost $T^k(x)$, $k \in \mathbb{N}$. Z hlediska pravděpodobnostního

zkoumání vystupují do popředí zejména nejrůznější časové průměry podél trajektorií. Označme proto pro $x \in \Omega$, $m \leq n$, měřitelnou $f : \Omega \to \mathbb{R}$ následující sumy a průměry:

$$S_{[m,n)}(f,x,T) := \sum_{k=m}^{n-1} f(T^k x), \qquad \hat{S}_{[m,n)}(f,x,T) := \frac{1}{n-m} S_{[m,n)}(f,x,T)$$
$$S_n(f,x,T) := S_{[0,n)}(f,x,T), \qquad \hat{S}_n(f,x,T) := \hat{S}_{[0,n)}(f,x,T).$$

O časových průměrech budeme mluvit v případě $\hat{S}_{[m,n)}(f,x,T)$, ale také v limitním případě, který značíme následovně:

$$\hat{S}(f,x,T) := \lim_{n\to\infty} \hat{S}_n(f,x,T).$$

Pokud bude $T$ jasné z kontextu, budeme ho často vypouštět a psát jen $S(f,x)$ namísto $S(f,x,T)$, atp.

Uvědomme si také, že notaci $S_{[m,n)}$ atp. lze zastoupit pomocí jednodušší $S_n$ následujícím způsobem:

$$S_{[m,n)}(f,x,T) := S_n(f,T^m x,T), \qquad \hat{S}_{[m,n)}(f,x,T) := \hat{S}_n(f,T^m x,T)$$

Zásadní větou, týkající se pravděpodobnostních dynamických systémů je ergodická věta. Ta má různé podoby a my zde uvedeme bez důkazu její formu, která se týká chování ve smyslu "skoro jistě". My ji uvedeme bez použití právě zavedené notace.

**Věta 3** (Birkhoffova ergodická věta)**.** *Nechť $(\Omega, \mathcal{A}, \mu, T)$ je invariantní pravděpodobnostní dynamický systém, $f : \Omega \to \mathbb{R}$ je $\mu$-integrovatelná. Pak existuje $\mu$-integrovatelná funkce $f^* : \Omega \to \mathbb{R}$ taková, že $f^*(T(x)) = f^*(x)$ s.j. a*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = f^*(x) s.j., \qquad \int f^*(x)d\mu = \int f(x)d\mu.$$

Ergodická věta tedy ukazuje, že pro většinu prvků $x \in \Omega$ se časové průměry hodnot $f$ ustalují, mají limitu. To je hlavní přínos ergodické věty a tento fakt má netriviální důkaz. To, že $f^*$ je invariantní skoro všude vyplývá již vcelku přímo z faktu, že $f^*$ je limitou časových průměrů. Ve speciálním případě, kdy je $f$ omezená, se průměr pro prvních $n$ iterací pro $x$ a pro $f(x)$ liší o $|f(x) - f(T^{n+1}(x))|/n$, což jde jistě k nule. Uveďme ještě tuto větu s použitím naší notace.

**Věta 4** (Birkhoffova ergodická věta, formulace II)**.** *Nechť $(\Omega, \mathcal{A}, \mu, T)$ je invariantní pravděpodobnostní dynamický systém, $f : \Omega \to \mathbb{R}$ je $\mu$-integrovatelná. Pak je časový průměr $\hat{S}(f,x)$ skoro všude dobře definovaný a platí*

$$\hat{S}(f,(T(x))) = \hat{S}(f,x) \text{ s.j.}, \qquad \int \hat{S}(f,x)d\mu = \int f(x)d\mu.$$

Ergodická věta je podobná zákonu velkých čísel, kde se říká, že limita časových průměrů konverguje skoro jistě ke střední hodnotě. Časový průměr $\hat{S}(f,x)$, jako funkce $x$, je tedy v takovém případě skoro jistě konstantní. Tuto sílu v případě ergodické věty pro obecný invariantní systém nemáme.

Na rozdíl od zákona velkých čísel $\hat{S}(f,x)$ limita může záležet na $x$. Jednoduše nemusí platit, že limitou časových průměrů je střední hodnota reálné veličiny $f$. Pokud bychom ovšem přidali konstantnost $\hat{S}(f,x)$ jako předpoklad, pak už by z druhé části ergodické věty plynulo, že tato konstanta se musí rovnat $\mathbb{E}(f)$. To nás vede k definici ergodického systému a následujícímu důsledku ergodické věty, který již zcela odpovídá zákonu velkých čísel.

**Definice 8.** *Dynamický systém $(\Omega, \mathcal{A}, \mu, T)$ je ergodický, pokud pro každou integrovatelnou $f : \Omega \to \mathbb{R}$ je $\hat{S}(f,x)$ konstantní skoro všude.*

**Důsledek 3.** *Invariantní dynamický systém $(\Omega, \mathcal{A}, \mu, T)$ je ergodický právě tehdy když pro každou integrovatelnou $f : \Omega \to \mathbb{R}$ platí:*

$$\hat{S}(f,x) = \mathbb{E}(f) \text{ s.j.}$$

Dále budeme potřebovat ekvivalentní charakteristiku ergodicity. Nejprve uvedeme potřebné značení a jednoduchá pozorování. První obecné tvrzení z teorie pravděpodobnosti uvádíme bez důkazu.

**Lemma 12.** *Pro integrovatelné funkce $f, g : \Omega \to \mathbb{R}$ na pravděpodobnostním prostoru platí:*

$$f \leq g \text{ s.j.} \quad \& \quad \int f \geq \int g \qquad \Rightarrow \qquad f = g \text{ s.j.}$$

*Pro měřitelnou funkci $f$ platí, že je skoro všude konstantní, právě tehdy když pro všechna $a, b \in \mathbb{R}$, $a < b$, platí*

$$\mu(f^{-1}(a,b)) \in \{0,1\}.$$

*Pro měřitelnou funkci $f$ platí, že je skoro všude konstantní, právě tehdy když pro všechna $b \in \mathbb{R}$ platí*

$$\mu(f^{-1}(-\infty,b)) \in \{0,1\}.$$

Pro posloupnost množin $U_i$, $i \in \mathbb{N}$ definujme

$$\liminf_{i\to\infty} U_i = \bigcup_{n=0}^{\infty} \bigcap_{i=n}^{\infty} U_i, \qquad \limsup_{i\to\infty} U_i = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} U_i.$$

První množina obsahuje prvky, které náleží do všech množin $U_i$ z posloupnosti, až na konečně mnoho, druhá množina obsahuje ty prvky, které náleží do nekonečně mnoha množin z posloupnosti $U_i$. Je tedy vidět, že druhá obsahuje první.

Pro nás budou zajímavé množiny

$$\liminf_{i\to\infty} T^{-i}U, \qquad \limsup_{i\to\infty} T^{-i}U.$$

Obě dvě limitní množiny jsou z definice invariantní.

**Lemma 13.** *Mějme invariantní dynamický systém $(\Omega, \mathcal{A}, \mu, T)$, $U \in \mathcal{A}$. Pokud je $U$ skoro sub-invariantní nebo skoro super-invariantní, pak je skoro invariantní, t.j.*

$$\big(\mu(T^{-1}U \setminus U) = 0 \text{ nebo } \mu(U \setminus T^{-1}U) = 0\big) \qquad \Rightarrow \qquad \mu(T^{-1}U \triangle U) = 0.$$

*Pokud je množina $U$ skoro invariantní, pak má stejnou míru jako invariantní množina* $\liminf_{i\to\infty} T^{-i}U$.

*Důkaz.* První část dokažme obecněji, buď $U, V \in \mathcal{A}$, $\mu(U) = \mu(V)$. Pak

$$\mu(U \setminus V) = \mu(U) - \mu(U \cap V) = \mu(U) - (\mu(V) - \mu(V \setminus U)) = \mu(V \setminus U).$$

Tedy je-li míra jednoho rozdílu nula, pak je i míra druhého rozdílu nulová.

Pokud je množina $U$ skoro-invariantní, pak

$$\begin{aligned}\mu(\bigcup_{k\geq n} T^{-k}(U)) &\leq \mu(T^{-n}U) + \sum_{k>n} \mu(T^{-k}U \setminus T^{-(k-1)}U) \\ &\leq \mu(T^{-n}U) + \sum_{k>n} \mu(T^{-1}U \setminus U) = \mu(T^{-n}U).\end{aligned}$$

Jelikož je $T^{-n}U$ podmnožinou daného sjednocení, musí být míry totožné. Vzhledem ke spojitosti pravděpodobnosti platí:

$$\mu(\liminf_{i\to\infty} T^{-i}U) = \lim_{n\to\infty} \mu(\bigcup_{k=n}^{\infty} T^{-k}U) = \lim_{n\to\infty} \mu(T^{-n}U) = \mu(U).$$

□

**Tvrzení 12** (Charakterizace ergodicity I)**.** *Pro invariantní dynamický systém $(\Omega, \mathcal{A}, \mu, T)$ jsou následující podmínky ekvivalentní:*

1. *systém je ergodický*
2. *Každá invariantní měřitelná množina má míru nula nebo jedna, t.j.*

$$U \in \mathcal{A} \quad \& \quad U = T^{-1}(U) \qquad \Rightarrow \qquad \mu(U) \in \{0, 1\}.$$

3. *Každá skoro-všude sub-invariantní měřitelná funkce je skoro všude konstantní.*

$$f(T(x)) \leq f(x) \text{ } s.j. \qquad \Rightarrow \qquad \exists c \in \mathbb{R}, f(x) = c \text{ } s.j.$$

*Důkaz.* Indikátor $1_U$ invariantní měřitelné množiny, je invariantní integrovatelnou funkcí. V invariantním ergodickém systému je tedy roven skoro všude konstantě. Je tedy buď skoro všude roven jedné, nebo nule. Ovšem střední hodnota indikátoru, alias míra $U$ pak musí být rovna této konstantě - tedy 0, nebo 1. Tím jsme dokázali implikaci (1) $\Rightarrow$ (2).

Předpokládejme nyní platnost (2) a nechť vezměme nyní skoro všude sub-invariantní měřitelnou funkci $f$. Pro $d \in \mathbb{R}$ uvažme množinu $V_d$ těch $x \in \Omega$, pro které $f(x) < d$. Pokud je nyní $f(T(x)) \leq f(x)$, platí také $T(x) \in V_d$. Neboli $V_d \setminus T^{-1}V_b$ je podmnožinou množiny, kde funkce $f$ není sub-invariantní. Tato množina má míru nula a proto je $V_d$ skoro super-invariantní. Potom ovšem, dle předchozího lemmatu existuje invariantní množina stejné míry. Z platnosti podmínky (2) dostáváme okamžitě, že $\mu(V_d)$ je nula, nebo jedna. Toto platí pro jakékoliv $d \in \mathbb{R}$. Splnili jsme tedy kritérium skoro konstantnosti pro $f$.

Předpokládejme platnost podmínky (3). Mějme integrovatelnou funkci $f : \Omega \to \mathbb{R}$. Z ergodické věty plyne, že je $\hat{S}(f, x)$ integrovatelná a skoro všude invariantní. Z podmínky (3) dostáváme, že je tedy skoro všude konstantní. □

Další důležitá charakterizace bude velmi užitečná při ověřování ergodicity.

**Tvrzení 13** (Charakterizace ergodicity II)**.** *Invariantní dynamický systém* $(\Omega, \mathcal{A}, \mu, T)$ *je ergodický právě tehdy když pro každé dvě množiny* $U, V \in \mathcal{A}$ *platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U \cap V) = \mu(U)\mu(V).$$

*Důkaz.* Pokud je systém ergodický, platí

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U \cap V) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \int 1_{T^{-k}U}(x) 1_V(x) d\mu(x) \\ &= \int \left( \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{T^{-k}U}(x) \right) 1_V(x) d\mu(x) \\ &= \int \mu(U) 1_V(x) d\mu(x) = \mu(U)\mu(V).\end{aligned}$$

Opačně, je-li $U \in \mathcal{A}$ invariantní, dostáváme

$$\mu(U) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U \cap U) = \mu(U)\mu(U).$$

Z toho plyne $\mu(U) \in \{0, 1\}$. □

**Tvrzení 14** (Charakterizace ergodicity III)**.** *Buď* $(\Omega, \mathcal{A}, \mu, T)$ *invariantní dynamický systém,* $\mathcal{A}'$ *systém množin, který generuje* $\mathcal{A}$*, který je uzavřený na průniky a každá jeho množina může být doplněna dalšími množinami z* $\mathcal{A}'$ *tak, že dohromady tvoří konečný rozklad* $\Omega$*. Pak je systém ergodický právě tehdy když pro každé dvě množiny* $U, V \in \mathcal{A}'$ *platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U \cap V) = \mu(U)\mu(V).$$

*Důkaz.* Přímá implikace plyne z předchozího tvrzení.

K důkazu zpětné opět použijeme kritérium pomocí invariantní množiny. Nejprve si ovšem uvědomme, že máme-li $U$ a $V$ ve formě disjunktních sjednocení množin z $\mathcal{A}'$, pak pro ně také platí rovnice z tvrzení. Konkrétně, je-li $U = \bigcup_{i=1}^{\ell} U_i$, $V = \bigcup_{j=1}^{m} V_j$, pro vzájemně disjunktní množiny $U_i$, $i \le \ell$ a pro vzájemně disjunktní množiny $V_j$, $j \le m$, pak Navíc

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U \cap V) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i=1}^{\ell} \sum_{j=1}^{m} \mu(T^{-k}U_i \cap V_j) \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{m} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U_i \cap V_j) \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{m} \mu(U_i)\mu(V_j) \\ &= \sum_{i=1}^{\ell} \mu(U_i) \sum_{j=1}^{m} \mu(V_j) = \mu(U)\mu(V).\end{aligned}$$

Připomeňme standardní fakta z teorie míry. Označíme-li množinu všech konečných disjunktních sjednocení množin z $\mathcal{A}'$ jako $\mathcal{A}''$, dostaneme algebru množin, t.j. soubor množin uzavřený na konečná sjednocení (nejen ta disjunktní), průniky a doplňky. Uvažme nyní rozšíření tohoto systému množin o všechny množiny $U \in \mathcal{A}$, pro které je pravda, že

$$\forall \varepsilon > 0, \qquad \exists U_\varepsilon \in \mathcal{A}'' : \mu(U \triangle U_\varepsilon) < \varepsilon.$$

Dá se lehce nahlédnout, že je toto rozšíření $\sigma$-algebrou, tedy obsahuje celé $\mathcal{A}$.

Máme-li tedy invariantní množinu $U$ a $\varepsilon > 0$, existuje $U_\varepsilon \in \mathcal{A}''$ taková, že $\mu(U \triangle U_\varepsilon) < \varepsilon$. Neboli

$$\begin{aligned} |\mu(U) - \mu(T^{-k}U_\varepsilon \cap U_\varepsilon)| &= |\mu(T^{-k}U \cap U) - \mu(T^{-k}U_\varepsilon \cap U_\varepsilon)| \\ &\leq \mu((T^{-k}U \cap U) \triangle (T^{-k}U_\varepsilon \cap U_\varepsilon)) \\ &\leq \mu(T^{-k}U \triangle T^{-k}U_\varepsilon) + \mu(U \triangle U_\varepsilon) \leq 2\varepsilon. \end{aligned}$$

Tedy

$$|\mu(U) - \mu(U_\varepsilon)\mu(U_\varepsilon)| = |\mu(U) - \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}U_\varepsilon \cap U_\varepsilon)| \leq 2\varepsilon.$$

Podobně

$$|\mu(U)\mu(U) - \mu(U_\varepsilon)\mu(U_\varepsilon)| \leq |\mu(U) - \mu(U_\varepsilon)|(\mu(U) + \mu(U_\varepsilon)) \leq 2\varepsilon.$$

Platí tedy, že $|\mu(U) - \mu(U)\mu(U)| \leq 4\varepsilon$. Toto ovšem platí pro libovolné $\varepsilon > 0$. Proto $\mu(U) = \mu(U)\mu(U)$ a $\mu(U) \in \{0, 1\}$. Systém je tedy ergodický. □

Bohužel až toto tvrzení umožňuje dokázat ergodicitu pro Bernoulliovské a Markovovy stacionární systémy. K tomu budeme používat následující důsledek.

**Tvrzení 15.** *Invariantní symbolický systém $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu, \sigma)$ je ergodický, právě tehdy když pro každé dvě slova $u, v \in A^*$ platí*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\sigma^{-k}[u] \cap [v]) = \mu([u])\mu([v]).$$

*Důkaz.* Cylindry jsou uzavřené na průniky, každý se dá doplnit dalšími cylindry slov stejné délky do rozkladu $A^{\mathbb{N}}$ a zároveň generují $\mathcal{B}(A^{\mathbb{N}})$. Lze tedy aplikovat předchozí tvrzení. □

**Důsledek 4.** *Invariantní symbolické systémy odpovídají stacionárním procesům. Definice ergodicity si odpovídají.*

**Příklad 6.** *Dokažte, že je Bernoulliovský posun ergodický.*

**Příklad 7.** *Dokažte, že je stacionární aperiodický nerozložitelný Markovský posun ergodický. Využijte toho, že pro něj existuje limitní (jedno-dimenzionální) rozdělení a jediné stacionární a tato rozdělení jsou si rovna, neboli pro každé $a \in A$,*

$$\lim_{n\to\infty} \mu(\sigma^{-n}[a]) = \pi(a),$$

*kde $\pi = (\pi(a))_{a\in A}$ je stacionární rozdělení. Zároveň je pro stacionární Markovský proces počáteční rozdělení rovno stacionárnímu.*

**Příklad 8.** *Dokažte, že je stacionární nerozložitelný Markovský posun ergodický. Využijte toho, že pro něj existuje jediné stacionární rozdělení, které je rovno časovému průměru jedno-dimenzionálních rozdělení, neboli pro každé $a \in A$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\sigma^{-k}[a]) = \pi(a),$$

*kde $\pi = (\pi(a))_{a\in A}$ je stacionární rozdělení. Zároveň je pro stacionární Markovský proces počáteční rozdělení rovno stacionárnímu.*

# 8 Asymptoticky rovnoměrné rozdělení

V této části našeho textu se budeme zabývat bohatostí jazyka pro daný invariantní ergodický symbolický systém, který zároveň budeme chápat jako prostor realizací náhodného procesu. Máme tedy invariantní ergodický symbolický systém $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu, \sigma)$ a náhodný proces $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ s rozdělením $P_{\mathbb{X}} = \mu$ (ten musí být pak také ergodický a stacionární).

Pro rychlejší zápis zaveďme notaci $h(\mu) = H(\mathbb{X})$:

$$h_n(x) = \mathcal{I}([x_{[0,n)}]) = -\log \mu([x_{[0,n)}]), \qquad H_n(\mu) = H(X_{[0,n)}) = \mathbb{E}(h_n)$$

Uvědomme si, že

$$\mathcal{I}_{X_{[0,n)}}(\omega) = h_n(\mathbb{X}(\omega)).$$

a že $\frac{1}{n}H_n(\mu)$ konverguje k $h(\mu)$.

Jazykem symbolického systému $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu, \sigma)$ se rozumí

$$A^*_\mu = \{u \in A^* \mid \mu([u]) > 0\}.$$

Množina takových slov je neprázdná a každé slovo z $A^*_\mu$ má prodloužení, které také náleží jazyka. To je jednoduchý důsledek toho, že cylindry $[ua]$ tvoří disjunktní rozklad cylindru $[u]$. Měl-li tedy cylindr $[u]$ nenulovou pravděpodobnost, musí mít nenulovou pravděpodobnost také cylindr $[ua]$ pro nějaké $A$.

Obraťme nyní svou pozornost na slova stejné délky, tedy na množiny

$$A^n_\mu = A^n \cap A^*_\mu,$$

$n \in \mathbb{N}$. Tyto množiny dobře popisují takzvaný nosič míry $\mu$, značený jako $s(\mu)$, což je průnik všech uzavřených množin plné míry. Zde je možné psát:

$$s(\mu) = \bigcap_{n\in\mathbb{N}} \bigcup_{u\in A^n_\mu} [u].$$

Nás bude zajímat, jak bohaté jsou tyto množiny a množiny odvozené. Pro symbolický systém uvažujme následující čtyři případy:

- Symbolický systém je Bernoulliovský na abecedě $(0, 1, 2)$ s počátečním vektorem $(1/3, 1/3, 1/3)$,
- Symbolický systém je Bernoulliovský na abecedě $(0, 1, 2)$ s počátečním vektorem $(1/2, 1/2, 0)$,

- Symbolický systém je Bernoulliovský na abecedě $(0,1,2)$ s počátečním vektorem $(49/100, 49/100, 2/100)$,
- Symbolický systém je Markovský na abecedě $(0,1,2)$ s počátečním vektorem $(1/3,1/3,1/3)$ a pravděpodobností přechodu $M_{ab} = \frac{1}{2}$, pokud $a \neq b$, jinak $M_{ab} = 0$.

V prvním a třetím případě je $A_\mu^n = A^n$. Míra $\mu$ tedy nijak nezmenšila prostor možných realizací. V druhém případě je $A_\mu^n = \{0,1\}^n \neq A^n$. Ve čtvrtém případě lze nahlédnout, že $A_\mu^n$ sestává ze slov, ve kterých nejsou žádné dva po sobě jdoucí symboly stejné. V případech 1,2 a 4 mají všechna slova z $A_\mu^n$ stejnou pravděpodobnost. V těchto případech je tedy také informační obsah $\mathcal{I}_{X_{[0,n)}}$ všude konstantní a roven $H(X_{[0,n)})$. Triviálně pak platí $x \in A^{\mathbb{N}}$ platí

$$\lim_{n\to\infty} \frac{1}{n} h_n(x) = h(\mu), \qquad x \in A^{\mathbb{N}}.$$

V řeči procesu,

$$\lim_{n\to\infty} \frac{1}{n} \mathcal{I}_{X_{[0,n)}} = h(\mu) \ s.j.$$

**Příklad 9.** *Spočti pravděpodobnost slova z $A_\mu^n$ ve čtvrtém případě.*

Trochu jiná je situace ve třetím případě, kde $A_\mu^n = A^n$, ale kde posloupnosti dvojek jsou daleko méně pravděpodobné, než posloupnosti nul, či jedniček. Dokonce, ani po zlogaritmování pravděpodobnosti a podělení délkou slova se tento rozdíl nebude blížit nule, neboť poměr mezi pravděpodobnostmi diverguje do nekonečna exponenciálně rychle. Konkrétně:

$$\frac{\mu[1^n]}{\mu[0^n]} = \left(\frac{49}{2}\right)^n, \qquad -\frac{1}{n}(\log \mu[0^n] - \log \mu[1^n]) = \log(49) - \log 2.$$

Zároveň

Na druhou stranu, se jedná o i.i.d. proces a díky zákonu velkých čísel platí, že ve skoro všech posloupnostech $\mathbb{X}(\omega)$ (v $\mu$-skoro všech posloupnostech $x \in A^{\mathbb{N}}$) se frekvence výskytů jednotlivých písmen limitně blíží jejich pravděpodobnosti, t.j.

$$\mathbb{P}\left(\bigcap_{a\in A} \left\{\omega \mid \frac{1}{n}\sum_{i=0}^{n-1} 1_{\{a\}}(X_i(\omega)) \to \mu([a])\right\}\right) = 1,$$

jinými slovy

$$\mu\left(\bigcap_{a\in A} \left\{x \in A^\infty \mid \frac{1}{n}\sum_{i=0}^{n-1} 1_{[a]}(\sigma^i x) \to \mu([a])\right\}\right) = 1.$$

Z toho také plyne, že pro každé $\varepsilon > 0$:

$$\mu\left(\bigcap_{a\in A} \left\{\omega \mid \left|\frac{1}{n}\sum_{i=0}^{n-1} 1_{[a]}(\sigma^i x) - \mu([a])\right| < \varepsilon\right\}\right) \to 1.$$

Ovšem

$$\mu([x_{[0,n)}]) = \prod_{i=0}^{n} \mu[x_i] = \prod_{a\in A} (\mu[a])^{\sum_{i=0}^n 1_{[a]}(\sigma^i x)}$$

a

$$\frac{1}{n} h_n(x) = -\frac{1}{n}\log \mu([x_{[0,n)}]) = -\sum_{a\in A} \frac{1}{n}\sum_{i=0}^{n} 1_{[a]}(\sigma^i x) \log \mu([a])$$

Tento výraz skoro jistě konverguje k $-\sum_{a\in A}\mu([a])\log\mu([a])$, což je entropie Bernoulliovského procesu. Tedy v tomto případě, nikoliv pro všechna $x$, ale pro skoro všechna $x \in A^{\mathbb{N}}$ platí

$$\lim_{n\to\infty}\frac{1}{n}h_n(x) = h(\mu).$$

Rozdíl mezi pravdivostí pro všechna $x$ a pro ($\mu$-)skoro všechna $x$ ilustruje následující cvičení.

**Příklad 10.** *Pro třetí případ dokažte, že pro počtem většinu slov z $A^n_\mu$, $-\frac{1}{n}\log(\mu([u]))$ nekonverguje k $h(\mu)$. (Hint - aritmetické proporci odpovídá proporce vůči rovnoměrnému rozdělení. To vznikne z rovnoměrného Bernoulliovského rozdělení, které má ale jinou entropii)*

Právě dokázané tvrzení pro Bernoulliovský případ platí ve skutečnosti pro jakýkoliv invariantní ergodický proces. Toto tvrzení, bude stěžejním tvrzením této kapitoly. Řekneme, že má symbolický systém (odpovídající náhodný proces) **asymptoticky rovnoměrné rozdělení**, pokud platí následující konvergence v pravděpodobnosti

$$\frac{1}{n}\left|h_n(x) - H_n(\mu)\right| \to 0, \qquad i.p.$$

V případě systému, který má dobře definovanou entropii, například v případě invariantního systému, to odpovídá podmínce

$$\frac{1}{n}h_n(x) \to h(\mu), \qquad i.p.$$

Řekneme, že má symbolický systém (odpovídající náhodný proces) **silně asymptoticky rovnoměrné rozdělení**, pokud platí konvergence skoro jistě, t.j.

$$\frac{1}{n}\left|h_n(x) - H_n(\mu)\right| \to 0, \qquad s.j.$$

Opět to v případě systému, který má dobře definovanou entropii, odpovídá podmínce

$$\frac{1}{n}h_n(x) \to h(\mu), \qquad s.j.$$

**Věta 5** (O asymptoticky rovnoměrném rozdělení)**.** *Každý invariantní ergodický symbolický systém má silně asymptoticky rovnoměrné rozdělení.*

Strategie důkazu bude nejprve ukázat, že horní a dolní limita $\frac{1}{n}h_n(x)$ se skoro všude rovná stejné konstantě. Zde použijeme kombinatorický rozbor dlouhých slov a ergodickou větu ($\delta$-pokrytí, $\delta$-rozklad, odhad možných složenin slov). Druhým krokem je dokázat, že se limita rovná entropii procesu. V této části důkazu je třeba nahlédnout fakt, že malá část slov délky $n$, malá z hlediska pravděpodobnosti, přispívá málo k entropii. V této úvaze vlastně ukazujeme, že jsou $\frac{1}{n}h_n(x)$ uniformně integrovatelné.

Jelikož je $[x_{[0,n+1)}]$ podmnožinou $\sigma^{-1}[(\sigma(x))_{[0,n)}]$, je $h_n(\sigma x)$ menší nebo rovno $h_{n+1}(x)$ a

$$\liminf_{n\to\infty}\frac{h_n(\sigma x)}{n} \le \liminf_{n\to\infty}\frac{h_{n+1}(x)}{n} = \liminf_{n\to\infty}\frac{h_n(x)}{n}.$$

Funkce $x \mapsto \liminf_{n\to\infty}\frac{h_n(x)}{n}$ je tedy subinvariantní. V ergodickém systému, který nás teď výhradně bude zajímat, je skoro všude rovna konstantě $\underline{h}$. Podobně je $\limsup_{n\to\infty}\frac{h_n(x)}{n}$ skoro všude rovna konstantě $\bar{h}$.

Tyto konstanty nazveme **dolní a horní entropií** a budeme chtít dokázat, že se rovnají entropii systému.

Definujme následující množiny slov nad abecedou $A$:

$$\begin{aligned} T_C^+ &= \{u \in A^* \mid \mu(u) < 2^{-|u|C}\} \\ T_C^- &= \{u \in A^* \mid \mu(u) > 2^{-|u|C}\} \\ T_C(\delta) &= T_{C-\delta}^+ \cap T_{C+\delta}^- \end{aligned}$$

Uvědomme si, že

$$x_{[0,n)} \in T^+(C) \text{ iff } \frac{h_n(x)}{n} > C.$$

Pro množinu slov $M \subset A^*$ zaveďme notaci $[M]$ která označuje sjednocení všech cylindrů pro slova z $M$, tedy

$$[M] = \bigcup \{[u] \mid u \in M\}.$$

Slova, a odpovídající cylindry, z množiny $T_{h(\mu)}(\delta)$ budeme nazývat typickými (z hlediska pravděpodobnosti). Zbylá slova a cylindry pak dělíme na příliš malá resp. velká, t.j. ty které náleží do $T_{h(\mu)+\delta}^+$ resp. $T_{h(\mu)-\delta}^-$. Důležitá nadmnožina typických slov je pak množina slov z $T_{h(\mu)+\delta}^-$, o které budeme označovat za velké (nikoliv příliš velké). Vše je samozřejmě relativní vůči parametru $\delta$.

Řekneme, že množina slov $M \subset A^*$ **slabě pokrývá** $(A^{\mathbb{N}}, \mu)$, pokud pro skoro všechna $x \in A^{\mathbb{N}}$, pro všechna $L$ existuje $n(x) > L$ takové, že $x_{[0,n(x))} \in M$, t.j.

$$\mu(\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} [M \cap A^k]) = 1.$$

Řekneme, že množina slov $M \subset A^*$ **pokrývá** $(A^{\mathbb{N}}, \mu)$, pokud

$$\lim_{n\to\infty} \mu([M \cap A^n]) = 1.$$

Dále řekneme, že $M$ **silně pokrývá** $(A^{\mathbb{N}}, \mu)$, pokud pro skoro všechna $x \in A^{\mathbb{N}}$, existuje $L$, tak že pro všechna $n > L$ platí $x_{[0,n)} \in M$.

$$\mu(\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} [M \cap A^k]) = 1.$$

Symbolický dynamický systém má tedy (silně) asymptoticky rovnoměrné rozložení, tehdy a jen tehdy, pokud pro každé $\delta > 0$, je $(A^{\mathbb{N}}, \mu)$ (silně) pokrytý slovy z $T_{h(\mu)}(\delta)$.

**Lemma 14.** *Pokud množina slov $M$ silně pokrývá $(A^{\mathbb{N}}, \mu)$, pak ho také pokrývá. Pokud množina slov $M$ pokrývá $(A^{\mathbb{N}}, \mu)$, pak ho také slabě pokrývá.*

*Důkaz.* Množiny $V_n = \bigcap_{k\geq n} [M \cap A^k]$, $n \in \mathbb{N}$, jsou rostoucí, vzhledem k inkluzi, proto je míra sjednocení limitou jejich měr. Dále

$$1 = \mu(\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} [M \cap A^k]) = \lim_{n\to\infty} \mu(\bigcap_{k\geq n} [M \cap A^k]) \leq \lim_{n\to\infty} \mu([M \cap A^n]).$$

Silné pokrývání tedy implikuje pokrývání.

Pokud $M$ pokrývá $(A^{\mathbb{N}}, \mu)$, pak pro každé $n$, $\bigcup_{k\geq n} [M \cap A^n]$ obsahuje množiny, jejichž míra jde k jedné. Sama tato množina má tedy míru jedna. Z toho již plyne, že $M$ slabě pokrývá $(A^{\mathbb{N}}, \mu)$. □

Zároveň pro $M$ zavedeme entropii takto:

$$\bar{H}(M) = \limsup_{n\to\infty} \frac{\log |M \cap A^n|}{n}.$$

Předpokládáme tedy, že počet slov v $M$ délky $n$ roste exponenciálně rychle. Entropie nám říká, jak rychlý je tento exponenciální růst.

Bohatost dynamického systému lze také chápat skrze to, jak velká množina slov je třeba, abychom náš dynamický systém pokryli. Definujeme tedy **dolní, resp. horní, pokrývací entropii**, $\underline{h}'$ a $\bar{h}'$, symbolického dynamického systému jako nejmenší možnou entropii množiny slov, která ho slabě, resp. silně, pokrývá, t.j.

$$\begin{aligned}
\underline{h}' &= \inf\{\bar{H}(M) \mid M \text{ slabě pokrývá } (A^{\mathbb{N}}, \mu)\} \\
h' &= \inf\{\bar{H}(M) \mid M \text{ pokrývá } (A^{\mathbb{N}}, \mu)\} \\
\bar{h}' &= \inf\{\bar{H}(M) \mid M \text{ silně pokrývá } (A^{\mathbb{N}}, \mu)\}
\end{aligned}$$

Diagonální argument pak ukazuje, že je možné infimum zaměnit za minimum.

Z definice dolní limity dostáváme, že pro každé $\delta > 0$ je $(A^{\mathbb{N}}, \mu)$ slabě pokrytý slovy z $T^-_{\underline{h}+\delta}$.

Z definice horní limity dostáváme, že pro každé $\delta > 0$ je $(A^{\mathbb{N}}, \mu)$ silně pokrytý slovy z $T^-_{\bar{h}+\delta}$.

Zároveň platí, že slova z $T^-_C$ délky $n$ generují vzájemně disjunktní cylindry míry nejméně $2^{-nC}$. Může jich tedy být maximálně $2^{nC}$. Dostáváme tedy,

$$\bar{H}(T^-_{\underline{h}+\delta}) \le \underline{h} + \delta, \qquad \bar{H}(T^-_{\bar{h}+\delta}) \le \bar{h} + \delta.$$

Z těchto pozorování plynou následující nerovnosti.

**Tvrzení 16.** *Pro ergodický systém, $\underline{h}' \le \underline{h}$ a $\bar{h}' \le \bar{h}$.*

Abychom navázali klasickou entropii $H(\mu)$ a všechny nově zmíněné entropie, potřebujeme konstruovat pokrytí jednotlivých dlouhých slov, pomocí kratších.

Posloupnost vzájemně disjunktních neprázdných intervalů $[m_i, n_i)$, $i \le k$, označíme za **$\delta$-rozklad** intervalu $[0, n)$, pokud ten obsahuje všechny intervaly z posloupnosti a platí $\sum_{i\le k}(n_i - m_i) > (1-\delta)n$.

Posloupnost vzájemně disjunktních neprázdných intervalů $[m_i, n_i)$, $i \le k$, označíme za **$M$-rozklad** intervalu slova $u$, pokud $u_{[m_i,n_i)} \in M$, $i \le k$ (mimojiné vyžadujeme $n_i \le |u|$, $i \le k$). Posloupnost intervalů nazveme $(M,\delta)$-rozkladem slova $u$, pokud je to $M$-rozklad slova $u$ a $\delta$-rozklad intervalu $[0, |u|)$.

Řekneme, že je slovo $u$ **$\delta$-pokryto** slovy z $M$, pokud existuje $(M,\delta)$-rozklad $u$. Množinu slov, které jsou $\delta$-pokryty slovy z $M$ označíme $M^{(\delta)}$.

**Lemma 15.** *Pro $n \in \mathbb{N}$, $0 < \delta \le 1/2$,*

$$\sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \le 2^{nH(\delta)},$$

*where $H(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$.*

*Důkaz.* Uvažujme výraz $x \log \delta) + (1-x)\log(1-\delta)$, $x \in [0,1]$. Výraz je váženým průměrem dvou hodnot, kde z předpokladu $\delta \leq 1/2$ plyne, že hodnota $\log \delta$ je menší nebo roven $\log(1-\delta)$. Proto, čím větší $x$, tím menší hodnota výrazu. Tedy pro $k \leq \delta n$,

$$\begin{aligned}\log \delta^k (1-\delta)^{n-k} &= k \log \delta + (n-k)\log(1-\delta) = n\left(\frac{k}{n}\log\delta + \left(1-\frac{k}{n}\right)\log(1-\delta)\right) \\ &\geq -nH(\delta).\end{aligned}$$

Dále

$$2^{-nH(\delta)} \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \leq \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \delta^k (1-\delta)^{n-k} \leq \sum_{k=0}^{n} \binom{n}{k} \delta^k (1-\delta)^{n-k} = (\delta + (1-\delta))^n = 1.$$

Z nerovnice již plyne požadovaný odhad. □

**Lemma 16.** *Mějme přirozené $L \geq 3$ a množinu slov $M$ takovou, že $|M \cap A^n| \leq 2^{nC}$, $n \in \mathbb{N}$, $|M \cap A^n| = 0$, $n \leq L$, $0 < \delta \leq 1/2$. Pak*

$$\bar{H}\left(M^{(\delta)}\right) \leq C + \delta \log |A| + H(L^{-1}) + H(\delta).$$

*Důkaz.* Ke každému slovu z $M^{(\delta)}$ přiřaďme jediný $\delta$-rozklad intervalu $[0,n)$, t.j. posloupnost intervalů $[m_i(u), n_i(u))$, $i \leq k(u)$, takový, že $u_{[m_i(u),n_i(u))} \in M$, $i \leq k(u)$.

Pro dané $n$ a daný pevný $\delta$-rozklad $[m_i, n_i)$, $i \leq k$, intervalu $[0,n)$, uvažujme slova z $M^{(\delta)} \cap A^n$, které mají tento rozklad jako svůj $(M,\delta)$-rozklad. Pro taková $u$ platí, že $u_{[m_i(u),n_i(u))} \in M$, $i \leq k(u)$. Slov z $M^{(\delta)} \cap A^n$ s daným rozkladem je tedy je maximálně tolik, kolika způsoby můžeme vyplnit pozice v každém z intervalů krát počet možností, kolika můžeme vyplnit zbylé pozice. Konkrétně je tento počet omezený konstantou

$$\prod_{i \leq k} 2^{(n_i - m_i)C} |A|^{\delta n} \leq 2^{\sum_{i \leq k}(n_i - m_i)} |A|^{\delta n} \leq 2^{nC} |A|^{\delta n}.$$

Zbývá ještě odhadnout, kolik existuje takových $\delta$-rozkladů intervalu $[0,n)$. Určit $\delta$ rozklad se dá například tak, že určíme, kde jsou startovní pozice jednotlivých intervalů a kde jsou mezery mezi nimi (tím jsou pak jasně vymezeny také konce intervalů). Jelikož je minimální délka intervalu $L$, umisťujeme nejvýše $\frac{n}{L} + 1$ startovních pozic. Takových možností je nejvýše

$$\sum_{k=0}^{\lfloor \frac{n}{L}+1 \rfloor} \binom{n}{k}.$$

Pro umístění mezer máme možností nejvýše

$$\sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k}.$$

Celkem tedy

$$\begin{aligned}\bar{H}\left(M^{\delta}\right) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \left( 2^{nC} |A|^{\delta n} \sum_{k=0}^{\lfloor \frac{n}{L}+1 \rfloor} \binom{n}{k} \sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} \right) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \left( 2^{nC} |A|^{\delta n} 2^{nH\left(\frac{\frac{n}{L}+1}{n}\right)} 2^{nH(\delta)} \right)\end{aligned}$$

Volili jsme $L \geq 3$, proto je pro dostatečně velké $n$, $\frac{n}{L} + 1$ menší než $n/2$. Proto jsme ve druhé nerovnosti mohli uplatnit na první součet kombinačních čísel příslušný odhad z předchozího lemmatu. Stejný odhad jsme uplatnili na druhý součet. □

**Lemma 17.** *Mějme množinu slov $M$ takovou, že slabě pokrývá ergodický systém $(A^{\mathbb{N}}, \mu)$. Potom pro každé $\delta > 0$, $M^{(\delta)}$ silně pokrývá $(A^{\mathbb{N}}, \mu)$.*

*Důkaz.* Fixujme $\delta > 0$. Z předpokladu dostáváme, že $\bigcup_{n\geq 1}[M \cap A^n]$ má plnou míru. Jistě najdeme $L \in \mathbb{N}$ takové, že

$$\mu\left[\bigcup_{n=1}^{L} M \cap A^n\right] \geq 1 - \delta/2.$$

Označme množinu slov z nerovnice jako $M_0$. Z ergodicity dostáváme, že pro skoro všechna $x \in A^{\mathbb{N}}$ platí, $\hat{S}(1_{[M_0]}, x) \geq 1-\delta/2$. Časový průměr $S(1_{[M_0]}, x)$ je vlastně frekvence návštěv bodů $(\sigma^i(x))$, $i \in \mathbb{N}$, v množině $[M_0]$. Ovšem $\sigma^i(x) \in [M_0]$ nastane právě tehdy když se nějaké slovo z $M_0$ vyskytuje v $x$ na pozici $i$. Uvědomme si, že v $\hat{S}(1_{[M_0]}, x)$ se zohledňují všechny výskyty slov z $M_0$, i ty, které se překrývají.

Uvažme nyní pro daný bod $x$ nepřekrývající se výskyty daných slov, které najdeme hladovým algoritmem. Tedy nejprve najdeme první pozici $r_1(x)$, kde se v $x$ vyskytuje slovo z $M_0$, které označíme $u_1(x)$. Pak hledáme první pozici větší než $r_1(x) + |u_1(x)|$, od které opět můžeme číst nějaké slovo z $M_0$. Danou pozici označíme $r_2(x)$ a slovo $u_2(x)$. Pokud již máme nalezené slovo $u_k(x)$ na pozici $r_k(x)$, hledáme výskyt slova z $M_0$ od pozice $r_k(x)+|u_k(x)|$, nalezené slovo značíme $u_{k+1}(x)$ a pozici $r_{k+1}(x)$. Tento postup opakujeme do nekonečna. Tímto způsobem jsme pokryli pozice

$$R(x) = \bigcup_{k\in\mathbb{N}} [r_k(x), r_k(x) + |u_k(x)|)$$

v $x \in A^{\mathbb{N}}$ slovy z $M_0$. Z definice hladového algoritmu pak platí, že $i \notin R(x)$ pouze tehdy, pokud se na $i$-té pozici nevyskytuje slovo z $M_0$.

Z toho pozorování a z faktu, že mají slova v $M_0$ uniformně omezenou délku, plyne že horní hustota $\mathbb{N} \setminus R(x)$ je nejvýše $\delta/2$, t.j.

$$\liminf_{n\to\infty} \frac{|[0,n) \cap R(x)|}{n} \geq 1 - \delta/2, s.j..$$

Navíc lze v předchozí nerovnosti zkrátit interval o konstantu, aniž by to změnilo hodnotu dolní limity. Platí tedy

$$\liminf_{n\to\infty} \frac{|[0,n-L) \cap R(x)|}{n} \geq 1 - \delta/2, s.j..$$

Pro $x$ splňující tuto nerovnost, existuje $n_0$ takové, že pro $n \geq n_0$ je
$|[0, n-L) \cap R(x)|$ větší než $n(1-\delta)$. Pro $n \geq n_0$, označme největší $k$, pro které $r_k(x) < n - L$. Jelikož jsou v $M_0$ délky slov maximálně $L$, je $[r_i(x), r_i(x) + |u_i(x)|)$ $M$-rozkladem slova $x_{[0,n)}$. Zároveň je

$$[0, n-L) \cap R(x) \subset \bigcup_{i\leq k} [r_i(x), r_i(x) + |u_i(x)|)$$

neboť levá strana sestává ze všech intervalů $[r_i(x), r_i(x)+|u_i(x)|)$, $i < k$ a části intervalu (možná úplné části) $[r_k(x), r_k(x)+|u_k(x)|)$. Platí tedy, že je daná posloupnost intervalů $(M, \delta)$-rozkladem $x_{[0,n)}$. Tedy $x_{[0,n)} \in M^{(\delta)}$. □

**Tvrzení 17.** *Pro invariantní ergodický proces platí, že dolní pokrývací entropie je rovna horní entropii, t.j. $\underline{h}' = \bar{h}$.*

*Důkaz.* Jistě platí $\underline{h}' \leq \bar{h}$. Vezměme libovolné $\varepsilon > 0$. Zvolme $\frac{1}{2} > \delta > 0$ a přirozené $L \geq 3$ takové, že

$$4\delta + \delta \log |A| + H(\delta) < \varepsilon/2, \qquad H(L^{-1}) < \varepsilon/2.$$

Z definice $\underline{h}'$ plyne, že existuje množina slov $M$ s entropií $\bar{H}(M) < \underline{h}' + \delta$, která slabě pokrývá daný invariantní ergodický systém. Existuje tedy $n_0$ takové, že pro všechny $n \geq n_0$, $|M \cap A^n| < 2^{n(\underline{h}'+2\delta)}$. Pokud nyní vyhodíme z množiny $M$ všechna slova délek menších než $n_0$, a zároveň všechna slova délek menších než $L$, dostaneme množinu $M_0$, která stále slabě pokrývá ergodický systém (vyhozením konečně mnoha slov se tato vlastnost nezmění), a platí pro ni $|M_0 \cap A^n| < 2^{n(\underline{h}'+2\delta)}$, $n \in \mathbb{N}$ a $|M_0 \cap A^n| = 0$, pro $n \leq L$.

Dostáváme, že $M_0^{(\delta)}$ silně pokrývá systém a zároveň

$$\bar{H}\left(M_0^{(\delta)}\right) \leq \underline{h}' + 2\delta + \delta \log |A| + H(L^{-1}) + H(\delta) \leq \underline{h}' + \varepsilon - 2\delta.$$

Tedy existuje $n_0$ takové, že pro všechny $n \geq n_0$, $|M_0^{(\delta)} \cap A^n| < 2^{n(\underline{h}'+\varepsilon-\delta)}$.

Z toho plyne, že

$$\mu\left[T^+_{\underline{h}'+\varepsilon} \cap M_0^{(\delta)} \cap A^n\right] \leq 2^{n(\underline{h}'+\varepsilon-\delta)} 2^{-n(\underline{h}'+\varepsilon)} = 2^{-n\delta},$$

pro $n \geq n_0$.

Odhad na pravé straně je sčítatelný. Z Borelova-Cantelliho principu pak plyne, že pro skoro všechna $x$ platí, že od jistého $n_1$, pro $n > n_1$, $x_{[0,n)}$ nenáleží do množiny $T^+_{\underline{h}'+\varepsilon} \cap M_0^{(\delta)}$. Podobně ovšem platí pro skoro všechna $x$, od jistého $n_2$, pro $n > n_2$, $x_{[0,n)}$ náleží do množiny $M_0^{(\delta)}$. Tyto dvě podmínky pak vymezují množinu (průnik) míry 1, kde pro každý bod $x$ z množiny existuje $n_3$ tak, že

$$x_{[0,n)} \notin T^+_{\underline{h}'+\varepsilon}, \qquad n \geq n_3.$$

Tedy

$$\limsup \frac{h_n(x)}{n} \leq \underline{h}' + \varepsilon, \quad s.j.$$

To platí pro jakékoliv $\varepsilon > 0$. Tedy $\bar{h} \leq \underline{h}'$. □

Z předchozích úvah vyplývalo, že $\underline{h}'$ a $\bar{h}$ jsou v rámci entropií $\underline{h}'$, $h'$, $\bar{h}'$, $\underline{h}$ a $\bar{h}$ minimem a maximem. Jelikož jsou si rovny, jsou si rovny všechny vyjmenované entropie. Tedy

$$\underline{h}' = \bar{h}' = h' = \underline{h} = \bar{h}.$$

Zbývá dokázat, že jsou všechny rovny entropii systému.

**Tvrzení 18.** *Pro invariantní ergodický systém* $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu, \sigma)$ *platí, že* $\underline{h} = \bar{h} = h(\mu)$.

*Důkaz.* Vezměme $\delta > 0$. Označme $T_n = T_{\underline{h}}(\delta) \cap A^n$, $T'_n = A^n \setminus T_n$. Z definice $\underline{h}$ a $\bar{h}$ plyne, že existuje $n_0$ takové, že pro všechna $n \geq n_0$, míra $[T_n]$ je minimálně $1 - \delta$.

Potom pro $n \geq n_0$ platí

$$H_n(\mu) \geq \sum_{u \in T_n, \mu([u])>0} \mu([u])(-\log \mu([u])) \geq (1-\delta)n(\underline{h} - \delta).$$

Připomeňme, že $h(\mu) = \lim_{n\to\infty} \frac{H_n(\mu)}{n}$. Tedy $h(\mu) \geq (1-\delta)(\underline{h} - \delta)$. Jelikož bylo $\delta > 0$ libovolné, dostáváme $h(\mu) \geq \underline{h}$.

Z druhé strany, pokud $\mu[T'_n] = 0$, dostáváme, že

$$H_n(\mu) = \sum_{u \in T_n, \mu([u])>0} \mu([u])(-\log \mu([u])) = \mu[T_n] n(\underline{h} + \delta) \leq n(\underline{h} + \delta).$$

Pokud $\mu[T'_n] > 0$, pak

$$\begin{aligned} H_n(\mu) &= \sum_{u \in T_n, \mu([u])>0} \mu([u])(-\log \mu([u])) + \sum_{u \in T'_n, \mu([u])>0} \mu([u])(-\log \mu([u])) \\ &\leq n(\underline{h} + \delta) + \mu[T'_n] \sum_{u \in T'_n, \mu([u])>0} \frac{\mu([u])}{\mu[T'_n]} \left( -\log \frac{\mu([u])}{\mu[T'_n]} \right) - \mu[T'_n] \log \mu[T'_n] \\ &\leq n(\underline{h} + \delta) + \delta n \log 2 + K, \end{aligned}$$

kde $K$ je maximum funkce $-x \log x$ na intervalu $(0, 1]$. Suma v druhém řádku nerovností je entropií normalizovaného rozdělení $\mu'(u) = \mu([u])/\mu[T'_n]$ na $T'_n$. Taková entropie je omezena logaritmem z počtu prvků, neboli $\log 2^n = n \log 2$. Dostáváme tedy, že

$$h(\mu) \leq \underline{h} + 2\delta.$$

Jelikož bylo $\delta > 0$ libovolné, je $h(\mu) \leq \underline{h}$. □

Tímto jsme dokázali rovnost všech zmíněných entropií, ale také větu 5.

# 9 Komprese

Mějme pevně dané konečné množiny $A$ a $B$, $B$ je konečné mohutnosti $D$.

Kód $f : A \to B^*$ indukuje pravděpodobnostní rozdělení $q^f$ na $A$:

$$q^f_a = \frac{D^{-|f(a)|}}{C(f)}, \qquad a \in A,$$

kde $C(f) = \sum_{a \in A} D^{-|f(a)|}$. Toto rozdělení hraje důležitou roli, viz následující věta.

**Tvrzení 19.** *Nechť $X$ je náhodná veličina s hodnotami v abecedě $A$, $f : A \to B^+$, $D = |B|$. Potom střední délka kódu splňuje:*

$$\mathbb{E}(|f(X)|) \log D = H(X) + D(P_X || q^f) - \log C(f).$$

**Důkaz:**

$$\begin{aligned} \mathbb{E}(|f(X)|) \log D - H(X) &= \sum_{a \in s(P_X)} P_X(a) \left( |f(a)| \log D + \log P_X(a) \right) \\ &= \sum_{a \in s(P_X)} P_X(a)(-\log(q^f_a C(f)) + \log P_X(a)) \\ &= -\log C(f) + D(P_X || q^f). \end{aligned}$$

Uvědomme si, že v poslední rovnosti hraje roli fakt, že $s(q^f) = A$, tedy že $s(P_X) \subset s(q^f)$. □

Pro prosté kódování do dané abecedy je důležitý počet slov délky nejvýše $k$, $d_k := \sum_{i=1}^{k} D^i$. Dále budeme používat hlavně odhad $D^k \leq d_k < D^{k+1}$, $k \in \mathbb{N}$.

**Tvrzení 20.** *Nechť $X$ je náhodná veličina s hodnotami v abecedě $A$, $f$ je prostý kód $f : A \to B^+$, $D = |B|$, $d_{k-1} < |A| \leq d_k$. Potom platí*

$$C(f) \leq k \leq 1 + \log_D |A|$$

*a*

$$\mathbb{E}(|f(X)|) \log D \geq H(X) - \log k \geq H(X) - \log(1 + \log_D |A|).$$

*Neboli*

$$\mathbb{E}(|f(X)|) \geq H_D(X) - \log_D(1 + \log_D |A|),$$

*kde $H_D(X)$ je hodnota entropie při použití logaritmu o základu $D$. Jinými slovy, je to entropie v $D$-itech.*

**Důkaz:** Zkonstruovat prostý kód $f'$ na $A$ s největším možným $C(f') = \sum_{a \in A} D^{[-|f'(a)|]}$ znamená vzít nejkratší možná kódová slova $f'(a)$, $a \in A$. Z omezení mohutnosti $|A|$ plyne, že mezi kódovými slovy budou všechna slova délky nejvýše $k - 1$ a několik slov délky nejvýše $k$. Pro takový kód, a tedy pro všechny prosté kódy, dostáváme omezení:

$$C(f') \leq \sum_{i=1}^{k} \sum_{u \in B^i} D^{-|u|} = k.$$

Zároveň platí, že $D^{k-1} < d_{k-1}$, pro $k \geq 2$. Z toho plyne $k \leq \log_D |A| + 1$. Pro $k = 1$, platí $k \leq \log_D |A| + 1$ triviálně. Zbytek tvrzení plyne z předchozího tvrzení a nezápornosti divergence. □

## 9.1 Asymptotické vlastnosti

**Definice 9.** *Pro náhodný proces $\mathbb{X}$ s hodnotami v abecedě $A$ a pro kód $f : A^+ \to B^+$ definujeme dolní a horní kompresní poměr takového kódu jako*

$$\underline{L}(\mathbb{X}, f) = \liminf_{n \to \infty} \frac{\mathbb{E}\left|f\left(X_{[0,n)}\right)\right|}{n}.$$

$$\bar{L}(\mathbb{X}, f) = \limsup_{n \to \infty} \frac{\mathbb{E}\left|f\left(X_{[0,n)}\right)\right|}{n}.$$

*Kompresní poměr je limita (pokud existuje):*

$$L(\mathbb{X}, f) = \lim_{n \to \infty} \frac{\mathbb{E}\left|f\left(X_{[0,n)}\right)\right|}{n}.$$

O kódu $f : A^+ \to B^+$ řekneme, že je slabě prostý, pokud každá jeho restrikce $f : A^n \to B^+$ je prostá.

**Věta 6.** *Buď $\mathbb{X}$ náhodný proces s výstupní abecedou $A$ a entropií $H(\mathbb{X})$. Nechť $f : A^+ \to B^+$ je slabě prostý kód. Potom dolní kompresní poměr je větší nebo roven entropii procesu v $D$-itech $H_D(\mathbb{X}) = \frac{H(\mathbb{X})}{\log D}$.*

*Důkaz.* Kód $f$ zúžený na $A^n$ je prostý, tedy platí

$$\mathbb{E}\left|f\left(X_{[0,n)}\right)\right| \log D \geq H(X_{[0,n)}) - \log(1 + \log_D |A^n|) = H(X_{[0,n)}) - \log(1 + n \log_D |A|).$$

Podělením přes $n$ a přechodem k dolní limitě dostáváme, že

$$\underline{L}(\mathbb{X}, f) \log D \geq H(\mathbb{X}) - \lim \frac{\log(1 + n \log_D |A|)}{n} = H(\mathbb{X}).$$

□

V předchozí části jsme zkoumali limitu posloupnosti středních hodnot, tedy jak se limitně chová průměrný poměr délky výstupního slova vůči délce vstupu. V této kapitole zesílíme naši pozornost a budeme se zabývat tím, jak se chová daný poměr bodově.

Kokrétně definujeme bodový dolní a horní kompresní poměr takového kódu pro kód $f : A^+ \to B^+$ a náhodný proces $\mathbb{X}$ v bodě $\omega \in \Omega$ předpisem

$$\underline{\ell}(\mathbb{X}, f)(\omega) = \liminf_{n\to\infty} \frac{\left|f\left(X_{[0,n)}(\omega)\right)\right|}{n}.$$

$$\bar{\ell}(\mathbb{X}, f)(\omega) = \limsup_{n\to\infty} \frac{\left|f\left(X_{[0,n)}(\omega)\right)\right|}{n}.$$

Bodový kompresní poměr je limita (pokud existuje):

$$\ell(\mathbb{X}, f)(\omega) = \lim_{n\to\infty} \frac{\left|f\left(X_{[0,n)}(\omega)\right)\right|}{n}.$$

Pokud budeme zkoumat střední hodnotu bodového kompresního poměru, budeme očekávat, že se rovná kompresnímu poměru definovanému dříve. Tak tomu bohužel vždy není, neboť nelze vždy prohodit pořadí střední hodnoty a limitního přechodu, aniž by to mělo vliv na výslednou hodnotu. Vzhledem k nezápornosti zkoumaných funkcí lze alespoň říci, že střední hodnota dolního bodového poměru je menší než dolní kompresní poměr - jedná se o důsledek Fatouova lemmatu.

**Tvrzení 21.** *Buď $\mathbb{X}$ náhodný proces s výstupní abecedou $A$ a buď $f : A^+ \to B^+$ slabě prostý kód. Potom platí*

$$\underline{\ell}(\mathbb{X}, f) \log D \geq \liminf_{n\to\infty} \frac{1}{n} \mathcal{I}_{X_{[0,n)}}, \ \text{s.j.}$$

*Důkaz.* Pro dané $n \in \mathbb{N}$, uvažujme množinu

$$V_n = \left\{\omega \in \Omega \mid \left|f\left(X_{[0,n)}(\omega)\right)\right| \log D < \mathcal{I}_{X_{[0,n)}}(\omega) - 3\log n\right\}$$

tedy množinu těch $\omega$, jejichž obraz $X_{[0,n)}$ kóduje $f$ lépe, než by naznačoval informační obsah $X_{[0,n)}$. Tato množina je volena tak, aby

$$\underline{\ell}(\mathbb{X}, f)(\omega) \log D < \liminf_{n\to\infty} \frac{1}{n} \mathcal{I}_{X_{[0,n)}}(\omega) \qquad \Longrightarrow \qquad \omega \in \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} V_k.$$

Neboli, pokud $\omega$ nesplňuje podmínku z tvrzení, pak musí patřit do nekonečně mnoha množin $V_n$. Naším cílem tedy je ukázat, že množina takových $\omega$ je nulová. Z definice $V_n$ plyne, že $V_n = \mathbb{X}^{-1}[W_n]$, kde

$$W_n = \left\{u \in A^n \mid |f(u)| < -\log_D P_{X_{[0,n)}}(u) - 3\log_D n\right\}.$$

Uvědomme si, že podmínka v definici $W_n$ je ekvivalentní s nerovností

$$0 < P_{X_{[0,n)}}(u) < \frac{1}{n^3} D^{-|f(u)|}.$$

Označme nyní restrikci $f$ na $s(P_{X_{[0,n)}}) \subset A^n$ jako $f_n$. Z definice $W_n$ plyne, že $W_n \subset s(P_{X_{[0,n)}})$ a

$$\begin{aligned}\mathbb{P}(V_n) &= \sum_{u\in W_n} P_{X_{[0,n)}}(u) < \frac{1}{n^3} \sum_{u\in W_n} D^{-|f_n(u)|} \leq \frac{1}{n^3} \sum_{u\in s(P_{X_{[0,n)}})} D^{-|f_n(u)|} \\ &= \frac{1}{n^3} C(f_n) \leq \frac{1 + \log_D |s(P_{X_{[0,n)}})|}{n^3} \leq \frac{1 + n\log_D |A|}{n^3}.\end{aligned}$$

Jelikož jde $(1+n\log_D|A|)/n$ k nule, je pro dostatečně velká $n$, $\mathbb{P}(V_n) \leq \frac{1}{n^2}$. To ovšem znamená, že

$$\sum_{n\in\mathbb{N}} \mathbb{P}(V_n) < \infty.$$

Z Borelova-Cantelliho principu nyní plyne, že skoro všechna $\omega$ náleží jen do konečně mnoha množin $V_n$. □

**Věta 7.** *Buď $\mathbb{X}$ náhodný proces s výstupní abecedou $A$, který má silně asymtoticky rovnoměrné rozložení a entropii $H(\mathbb{X})$ (například invariantní ergodický proces). Potom*

$$\underline{\ell}(\mathbb{X}, f) \geq \frac{H(\mathbb{X})}{\log D} = H_D(\mathbb{X}),\ s.j.$$

Podmínky na komprimovací kód $f : A^+ \to B^+$ by se daly, vzhledem k praktickému použití, ještě trochu zeslabit. Mohli bychom pouze vyžadovat, aby byl kód, pro dané $n$, definovaný jen na $s(P_{X_{[0,n)}})$, tedy jen pro slova, která se objevují s nenulovou pravděpodobností. Jen ty je totiž třeba v praxi umět zakódovat. Ovšem odhady na kompresní poměr jsou všechny odvozené z odhadu z Tvrzení 18 aplikované na restrikci kódu $f_n : A^n \to B^+$. Pokud bychom tuto restrikci ještě zúžili a uvažovali jen $f : s(P_{X_{[0,n)}}) \subset A^n \to B^+$, potřebný odhad by stále platil. Opravdu

$$\mathbb{E}\big|f\left(X_{[0,n)}\right)\big| \log D \geq H(X_{[0,n)}) - \log(1+\log_D|s(P_{X_{[0,n)}})|) \geq H(X_{[0,n)}) - \log(1+\log_D|A^n|).$$

Tedy i toto oslabení předpokladů vede ke stejnému závěru, že limitní bodový i nebodový kompresní poměr skoro jistě nepodběhne entropii invariantního ergodického procesu, který komprimujeme.

## 9.2 Efektivní kompresní kódy

V minulé kapitole jsme ukázali, že kompresní kód nemůže mít lepší asymptotický kompresní poměr než je entropie zdrojového ergodického procesu. V této části ukážeme naopak způsob, jak konstruovat kódy, které této hranice dosahují. Nejprve ukážeme, jak ušít takový kompresní kód procesu na míru. Jednotlivé kompresní kódy se tak pro různé vstupní procesy liší. Nakonec ukážeme konstrukci univerzálního kódu, tedy takového, který umí efektivně zkomprimovat jakýkoliv ergodický proces.

**Lemma 18.** *Nechť $X$ je náhodná veličina s hodnotami v abecedě $A$. Pak existuje prostý kód $f : A \to B^+$, takový, že*

$$|f(a)| = \lceil -\log_D P_X(a)\rceil, a \in s(P_X).$$

*Tedy*

$$|f(X)|\log D \leq \mathcal{I}_X + 1,\ s.j.\ , \qquad \mathbb{E}|f(X)| \leq H_D(X) + 1.$$

*Důkaz.* Uvažujme množinu $V_n = \{a \in A \mid \lceil -\log_D P_X(a)\rceil = n\}$. Z definice plyne, že $P_X(a) \geq D^{-n}$ pro každé $a \in V_n$. Proto $|V_n| \leq D^n \leq |B^n|$. Existuje tedy prosté zobrazení $f_n : V_n \to B^n$. Položme $f = \bigcup_{n\geq 1} f_n$. Jelikož jsou obory hodnot různých $f_n$ disjunktní, je kód $f$ prostý. Zároveň $s(P_x) = \bigcup_{n\geq 1} V_n$ a platí pro něj platí, $|f(a)| = \lceil -\log_D P_X(a)\rceil$, $a \in s(P_X)$. Zbytek tvrzení lemmatu je okamžitým důsledkem právě dokázaného. □

**Tvrzení 22.** *Nechť $\mathbb{X}$ je náhodný proces s hodnotami v abecedě $A$. Pak existuje slabě prostý kód $f : A^+ \to B^+$, takový, že*

$$\bar{\ell}(f, X)\log D \leq \limsup_{n\to\infty} \frac{1}{n}\mathcal{I}_{X_{[0,n)}},\ s.j.\ , \qquad \bar{L}(f, X) \leq \limsup_{n\to\infty} \frac{1}{n} H_D(X).$$

*Důkaz.* Pro každé $n$ definujeme $f : A^n \to B^+$ jako kód z předchozího lematu. Tím je definován kód $f : A^+ \to B^+$, který má prosté restrikce na jednotlivých $A^n$. Pokud nerovnosti v předchozím lematu podělíme přes $n$ a přejdeme k horní limitě, dostaneme důkaz tohoto tvrzení. □

**Důsledek 5.** *Nechť $\mathbb{X}$ je náhodný proces s hodnotami v abecedě $A$, který má silně asymptoticky rovnoměrné rozložení a entropii $H(\mathbb{X})$. Potom existuje slabě prostý kód $f : A^+ \to B^+$, takový, že*

$$l(f, X) = \frac{H(\mathbb{X})}{\log D}, \text{ s.j. }, \qquad L(f, X) = \frac{H_D(\mathbb{X})}{\log D}.$$

V tomto tvrzení jsme zkombinovali předchozí horní odhad pro horní limity a fakt, že v předchozí kapitole jsme dokázali stejné dolní odhady pro dolní limity.

Pokud bychom chtěli najít prostý, namísto jen slabě prostého kódu, je možné původní kód modifikovat tak, že mu předřadíme informaci o délce posílané zprávy. Pro jednoduchost nyní předpokládejme, že abeceda $B$ sestává z čísel 0 až $D-1$ (přeznačení abecedy nemá vliv na kompresní poměry, ani na entropie). Označme $\beta_D(n)$ $D$-ární zápis čísla $n$, tedy například $\beta_2(5) = 101$, či $\beta_3(15) = 120$. Podobně jako v binárním zápise ($D = 2$), platí pro $n \geq 1$

$$|\beta_D(n)| = \lfloor \log_D(n) \rfloor + 1 \leq \log_D(n) + 1.$$

Pokud bychom ale nějakému slabě prostému kódu $f$ předřadili tuto informaci, neboli vytvořili kód $g(u) = \beta_D(|u|)f(u)$, kód stále nemusí být prostý. Například, pokud by $f(0) = 00$ a $f(00) = 0$, pak by $g(0) = 100 = g(00)$. Jistotu ale získáme například tehdy, bude-li informace o délce prefixová.

Kód $f : A \to B^+$ nazveme **prefixovým** pokud platí, že $f(u)$ je prefixem $f(v)$ pouze v případě, že $u = v$. Uvědomme si, že prefixový kód je jistě prostý, tedy i slabě prostý.

Definujme nyní kódy $\gamma_i : \mathbb{N} \to B^+$ iterativně:

$$\gamma_0(n) = 1^n 0, \quad n \in \mathbb{N}$$

$$\gamma_{k+1}(n) = \gamma_k(|\beta_D(n)|)\beta_D(n), \quad n \in \mathbb{N}.$$

My budeme z těchto kódů potřebovat jen $\gamma_1$ a $\gamma_2$, kde

$$\gamma_1(n) = 1^{|\beta_D(n)|}0\beta_D(n), \qquad \gamma_2(n) = 1^{|\beta_D(|\beta_D(n)|)|}0\beta_D(|\beta_D(n)|)\beta_D(n).$$

**Lemma 19.** *Buď $f : A \to B^+$ prostý, $g : \mathbb{N} \to B^+$ prefixový. Pak je $f' : A \to B^+$, definovaný předpisem $f'(u) = g(|f(u)|)f(u)$, prefixový.*

*Kódy $\gamma_n$, $n \in \mathbb{N}$, jsou prefixové.*

*Důkaz.* Buď $f'(u)$ prefixem $f'(v)$. Potom $g(|f(u)|)$ a $g(|f(v)|)$ jsou oba prefixem stejného slova $f'(v)$. Proto musí být $g(|f(u)|)$ prefixem $g(|f(v)|)$ nebo naopak. V obou případech z prefixovosti plyne, že $|f(u)| = |f(v)|$, tedy i $g(|f(u)|) = g(|f(v)|)$, což je společný prefix slov $f'(u)$ a $f'(v)$. Aby bylo první prefixem druhého, musí být zbylá část $f(u)$ prefixem zbylé části $f(v)$. Tyto slova mají ovšem stejnou délku. Proto si musí být rovna. Z prostoty $f$ dostáváme, že $u = v$.

Prefixovost $\gamma_0$ je zřejmá. Z prefixovosti $\gamma_k$ a z první části lemmatu plyne prefixovost $\gamma_{k+1}$. □

**Lemma 20.** *Buď $f : A^+ \to B^+$ slabě prostý, $g : \mathbb{N} \to B^+$ prefixový. Pak je $f' : A^+ \to B^+$, definovaný předpisem $f'(u) = g(|u|)f(u)$, prostý.*

*Důkaz.* Buď $f'(u) = f'(v)$. Potom $g(|u|)$ a $g(|v|)$ jsou oba prefixem stejného slova. Proto musí být $g(|u|)$ prefixem $g(|v|)$ nebo naopak. V obou případech z prefixovosti plyne, že $|u| = |v|$. Ovšem $f$ je slabě prostý, tedy $u = v$. □

**Lemma 21.** *Pro kódy $\gamma_1, \gamma_2 : \mathbb{N} \to B^+$ jsou funkce $|\gamma_1|, |\gamma_2| : \mathbb{N} \to \mathbb{N}$ neklesající a a pro každé $n \geq 1$*

$$|\gamma_1(n)| \leq 2\log_D n + 3, \qquad \gamma_2(n) \leq \log_D n + 2\log_D(\log_D n + 1) + 4.$$

**Důkaz:** Omezení délek kódů vyplývají z konstrukce a z omezení pro délku kódu $\beta$ výše. □

Nyní už můžeme zlepšit předchozí věty do následujícího tvaru.

**Tvrzení 23.** *Nechť $\mathbb{X}$ je náhodný proces s hodnotami v abecedě $A$. Pak existuje prostý kód $f : A^+ \to B^+$, takový, že*

$$\bar{\ell}(f, X)\log D \leq \limsup_{n\to\infty} \frac{1}{n}\mathcal{I}_{X_{[0,n)}}, \text{ s.j. }, \qquad \bar{L}(f, X) \leq \limsup_{n\to\infty} \frac{1}{n} H_D(X).$$

*Důkaz.* Vezměme slabě prostý kód $f$, který splňuje podmínky tvrzení, který jistě existuje, viz Tvrzení 22. Uvažujme kód $f'$ definovaný předpisem $f'(u) = \gamma_1(|u|)f(u)$. Tento kód je dle předchozích tvrzení prostý a zároveň$|f'(u)| - |f(u)|$ je $o(|u|)$. Z toho plyne, že i pro $f'$ platí dané asymptotické odhady. □

**Důsledek 6.** *Nechť $\mathbb{X}$ je náhodný proces s hodnotami v abecedě $A$, který má silně asymptoticky rovnoměrné rozložení a entropii $H(\mathbb{X})$. Potom existuje prostý kód $f : A^+ \to B^+$, takový, že*

$$l(f, X) = \frac{H(\mathbb{X})}{\log D}, \text{ s.j. }, \qquad L(f, X) = \frac{H_D(\mathbb{X})}{\log D}.$$

Viděli jsme, že do konstrukce prostého kódu stačilo vhodně zakomponovat kód $\gamma_1$. S kódem $\gamma_2$ bychom pro velká $n$ dosáhli trochu lepších výsledků, ale v limitě by to vyšlo nastejno. V následující podkapitole bude ovšem použití $\gamma_2$ nutné. Méně úsporný kód $\gamma_1$ by tam nestačil.

## 9.3 Univerzální Lempel-Ziv kompresní kód

V této části představíme kód, který dokáže dobře zkomprimovat jakýkoliv ergodický zdroj.

Dané slovo $u \in B^*$ rozložíme na bloky proměnné délky tak, že začneme prázdným slovem a pokračujeme vždy nejkratším úsekem, který se mezi předcházejícími bloky nevyskytuje. Pokud po ukončení takového algoritmu zbude neprázdný koncový blok z $u$ přidáme ho do souboru bloků. Této reprezentaci říkáme **LZ-rozbor**. Například LZ-rozbor slova

$$u = 000011001110011001101110 0$$

je

$$R(u) = y = \lambda, 0, 00, 01, 1, 001, 11, 0011, 00110, 111, 00,$$

kde $\lambda$ značí prázdné slovo.

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $y^{(i)}$ | $\lambda$ | 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 10.** *LZ-Rozbor slova $u \in B^*$ je posloupnost slov $y = \left(y^{(i)}\right)_{i\leq C(u)}$, kde $y^{(0)} = \lambda$, $k_0 = 0$ a $y^{(i)} = u_{[k_{i-1},k_i)}$ pro $i > 0$. Zde , a pokud $k_i < |u|$ definujeme $k_{i+1} = \min J_i$, je-li množina*

$$J_i = \{j \in (k_i, |u|], \text{ s.j. } \forall m \leq i, u_{[k_i,j)} \neq y^{(m)}\}$$

*neprázdná, a $k_{i+1} = |u|$, je-li $J_i$ prázdná. Číslo $C(u)$ je nejmenší, pro které $k_{C(u)} = |u|$.*

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 $y^{(i)} = y^{(a_i)}c_i$.

Vezměme nyní prefixové kódování přirozených čísel $\gamma_2$ a zvolme pevně libovolný prefixový kód $f : A \to B^+$, který kóduje abecedu $A$. To lze udělat například tak, že zvolíme $n$ takové, že $D^n \geq |A|$ a za kódová slova vybereme příslušný počet slov délky $n$. Pokud je vstupní abeceda podmnožinou výstupní, můžeme za $f$ vzít identitu.

Lempelův-Zivův kód, dále jen LZ-kód, slova $u$ pak vypadá takto:

$$\mathrm{r}(u) = \gamma_2(a_1)f(c_1)\gamma_2(a_2)f(c_2)\dots\gamma_2(a_{C(u)})f(c_{C(u)}).$$

Pokud tedy kódujeme uvažované slovo $u = 000011001110011001101110 0$ do binární abecedy a $f$ je identita, dostaneme kód

$$\mathrm{r}(u) \quad = \quad 1010\mathbf{0}\,1011\mathbf{0}\,1011\mathbf{1}\,1010\mathbf{1}\,11010101\mathbf{1}\,110111001\mathbf{1}\,110111011\mathbf{1}\,1101111\mathbf{0}\,110111101\mathbf{1}\,1011\mathbf{0},$$

kde mezerou jsou odděleny kódy pro jednotlivé slova z LZ-rozboru, silně je zvýrazněn poslední symbol, který vždy odpovídá kódu $f(c_i)$.

**Lemma 22.** *LZ-kód je prostý.*

**Důkaz:** Buď $u, v \in A^+$, $\mathrm{r}(u) = \mathrm{r}(v)$. Nechť $y^{(i)}, a_i, c_i, k_i$, $i \leq C(u)$ jsou slova a koeficienty odvozené z LZ-rozboru $u$, $a'_i, c'_i, k'_i$, $i \leq C(u)$ jsou koeficienty odvozené z LZ-rozboru $v$. Dostáváme tedy, že $\gamma_2(a_1)$ i $\gamma_2(a'_1)$ jsou prefixem téhož slova, proto je jedno prefixem druhého. Z prefixovosti $\gamma_2$ dostáváme, že $a_1 = a'_1$. To tedy znamená, že $y^{(a_1)} = y^{(a'_1)}$ a také $k_1 = k'_1$. Pokud z $\mathrm{r}(u) = \mathrm{r}(v)$ odtrhneme společný prefix $\gamma_2(a_1)$, pak opět dostáváme, že $f(c_1)$ a $f(c_2)$ jsou prefixy téhož slova. Opět je tedy jedno prefixem druhého a z prefixovosti kódu $f$ dostáváme $c_1 = c'_1$ a tedy $y^{(1)} = y'^{(1)}$. Stejným způsobem pak dostaneme, že $a_2 = a'_2$, $k_2 = k'_2$ a $c_2 = c'_2$, $y^{(2)} = y'^{(2)}$. Takto můžeme postupovat až k rovnostem $a_\ell = a'_\ell$, $k_\ell = k'_\ell$ a $c_\ell = c'_\ell$, $y^{(\ell)} = y'^{(\ell)}$, kde $\ell$ je minimum z délek LZ-rozborů $C(u)$ a $C(v)$. Bez újmy na obecnosti předokládejme, že $\ell = C(u) \leq C(v)$. Ovšem kód $\gamma_2$ má vždy pozitivní délku, proto pokud by rozklad $v$ byl delší, byl by i LZ-kód $\mathrm{r}(v)$ vlastním rozšířením

$$\gamma_2(a_1)f(c_1)\gamma_2(a_2)f(c_2)\dots\gamma_2(a_\ell)f(c_\ell)$$

a nebyl by tak roven $\mathrm{r}(u)$. Tedy délky LZ-rozborů jsou stejné. Ovšem i slova v LZ-rozborech jsou stejné, ve stejném pořadí, proto jsou $u$ a $v$ shodná. □

Efektivita LZ-komprese, závisí na velikosti LZ-rozboru, neboli na $C(u)$ následujícím způsobem.

**Lemma 23.** *LZ-kód splňuje omezení*

$$|\mathrm{r}(u)| \leq C(u)\left(\log_D C(u) + 2\log_D(\log_D C(u) + 1) + 4 + K\right),$$

*kde $K$ je konstanta, která dominuje délky kódových slov kódu $|f|$.*

*Pro $x \in A^{\mathbb{N}}$,*

$$\limsup_{n\to\infty} \frac{|\mathrm{r}(x_{[0,n)})|}{n} \leq \limsup_{n\to\infty} \frac{C(x_{[0,n)})\log_D n}{n}.$$

*Důkaz.* Z konstrukce kódu vyplývá, že

$$|\mathrm{r}(u)| = \sum_{i=1}^{C(u)} \left(|\gamma_2(a_i)| + |f(c_i)|\right),$$

pro čísla $C(u)$, $a_i$ a $c_i$ odvozené z rozkladu slova $u$. Z definice plyne, že $a_i < C(u)$. Z monotónnosti logaritmu a z odhadu pro $\gamma_2$ z lematu 21 pak získáváme pro kladná $a_i$:

$$|\gamma_2(a_i)| \leq \log_D C(u) + 2\log_D(\log_D C(u) + 1) + 4.$$

Ovšem platnost stejného odhadu pro případ $a_i = 0$ lze ověřit pouhým dosazením.

Máme tedy uniformní odhad, který již zaručuje platnost nerovnosti z lemmatu. Druhou část tvrzení získáme jednoduchým limitním přechodem s využitím nerovnosti $|C(x_{[0,n)})| \leq n$ a toho, že iterovaný logaritmus je řádově menší než logaritmus neiterovaný. □

Ve zbytku kapitoly budeme dokazovat, že pro skoro všechna $x$, $C(x_{[0,n)})$ roste pomaleji než $(h+\varepsilon)n/\log n$, kde $h$ je entropie dynamického systému a $\varepsilon > 0$ je livolně malé.

Při této analýze budeme uvažovat větší množinu rozborů, takzvané 2-rozbory.

**Definice 11.** *Řekneme, že je konečná posloupnost $y = (y_i)_{i=1}^n$ neprázdných slov 2-rozborem slova $u \in A^*$, dále jen "rozborem" slova $u$, pokud je $u$ rovno konkatenaci slov z posloupnosti $v$ daném pořadí a pokud se jakékoliv slovo vyskytuje v posloupnosti maximálně dvakrát. Délkou rozboru rozumíme délku posloupnosti $y$ a značíme ji $|y|$.*

LZ-rozboru $y = (y_i)_{i=0}^{C(u)}$ slova $u$, odpovídá rozbor $y'$ slova $u$, který vznikne pouze zapomenutím slova $y_0 = \lambda$. Platí pak $|y'| = C(u)$. Tato vazba byla motivací, proč v rozboru povolit duplicitu slov (omezenou maximálně dvěma výskyty).

Množinu všech 2-rozborů slova $u$ označíme jako $R(u)$ a množinu všech 2-rozkladů všech slov délky $n$ jako $R(n)$. Dále označme maximální délku rozkladů z daných množin jako $L(u)$ a $L(n)$, t.j.

$$L(u) = \max\{|y| \mid y \in R(u)\}, \qquad L(n) = \max\{|y| \mid y \in R(n)\}.$$

Uvědomme si také důležitý vztah mezi průměrnou délkou slova z rozboru $y$ slova $u$:

$$\bar{y} = \frac{1}{|y|}\sum_{i=1}^{|y|}|y_i| = \frac{|u|}{|y|}.$$

Chceme-li tedy při důkazu efektivity, aby délka rozboru $x_{[0,n)}$ rostla pomaleji než $\frac{(h+\varepsilon)n}{\log n}$, požadujeme jinými slovy aby průměrná délka slov v rozboru $x_{[0,n)}$ rostla rychleji než $\frac{\log n}{h+\varepsilon}$. Tato myšlenka nás vede k tomu, zkoumat, jak velký kus daného slova délky $n$ pokrývají slova z rozboru délky menší než $\frac{\log n}{h+\varepsilon}$ a jak velkou část pokrývají slova delší. Stejně důležité bude jak velkou část z hlediska počtu pokrývají kratší a delší slova.

Z vět o pokrývací entropii víme, že entropie kvantifikuje jak bohatý systém slov máme, ze kterého můžeme vybírat. Zároveň je přirozené, že pokud v rozboru můžeme uplatnit jen slova z nějaké omezené množiny slov, například pouze určitá typická slova pro dané rozdělení procesu, jsme dříve nuceni použít také delší slova, jelikož krátkých slov je málo. Konkrétně o tom mluví následující lemma.

**Lemma 24.** *Mějme $C, \delta > 0$ a množinu slov $M \subset A^*$ takovou, že $|M \cap A^n| \leq 2^{Cn}$. Potom pro každé slovo $u \in A^+$ a každý jeho rozbor $y$, platí že slov délky nejvýše $\frac{\log|u|}{C+\delta}$ je v rozboru nejvýše*

$$\frac{2^{C+1}}{2^C - 1}|u|^{\frac{C}{C+\delta}}.$$

*Tato slova tedy souhrnně pokrývají nejvýše*

$$\frac{2^{C+1}}{C(2^C-1)}|u|^{-\frac{\delta}{C+\delta}} \cdot \log|u|$$

*pozic v $u$.*

*Důkaz.* Pro dané $k \geq 1$ platí, že slov délky nejvýše $k$ může být v rozboru maximálně dvojnásobek celkového počtu slov nejvýše $k$ z $M$, t.j. nejvýše

$$2\sum_{j=1}^{k} 2^{jC} = 2 \cdot \frac{2^{(k+1)C} - 1}{2^C - 1}.$$

Relativní část $u$, pokrytá slovy v rozboru $u$ délky nejvýše $\frac{\log|u|}{C+\delta}$ je tedy nejvýše

$$\frac{2}{|u|} \sum_{j=1}^{\lfloor \frac{\log|u|}{C+\delta} \rfloor} 2^{jC} = \frac{2}{|u|} \cdot \frac{2^{(\lfloor \frac{\log|u|}{C+\delta} \rfloor + 1)C} - 1}{2^C - 1} < \frac{2}{|u|} \cdot \frac{2^{(\frac{\log|u|}{C+\delta} + 1)C}}{2^C - 1} \leq \frac{2^{C+1}}{2^C - 1} |u|^{-\frac{\delta}{C+\delta}}.$$

Tvrzení o pokrytí vyplývá bezprostředně z omezení délky těchto slov. □