Úvod do teorie informace

Matematické základy komprese a digitální komunikace

Tomáš Kroupa
http://staff.utia.cas.cz/kroupa/

Ústav teorie informace a automatizace
AV ČR

2014

Obsah

Část I

Úvod

Teorie informace je matematická disciplína, která zkoumá

- možnosti komprese informace,
- metody rychlého a kvalitního přenosu informace.

Teorie informace je založena na **pravděpodobnostním modelu** zpráv a komunikace. Tato myšlenka umožnila **C. Shannonovi** v r. 1948 ukázat převratný fakt:

- zvyšování výkonu přenosového zařízení není jediná cesta k potlačení chyb při přenosu informace, neboť
- existuje určitá přenosová rychlost, od níž lze přenášet s libovolně malou pravděpodobností chyby.

Počátky teorie informace

Suppose we have a set of possible events whose probabilities of occurence are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or how uncertain we are of the outcome?

C.E. Shannon (1916–2001)

C. E. Shannon.
A mathematical theory of communication.
Bell System Tech. J., 27:379–423, 623–656, 1948.

Model komunikace podle Shannona

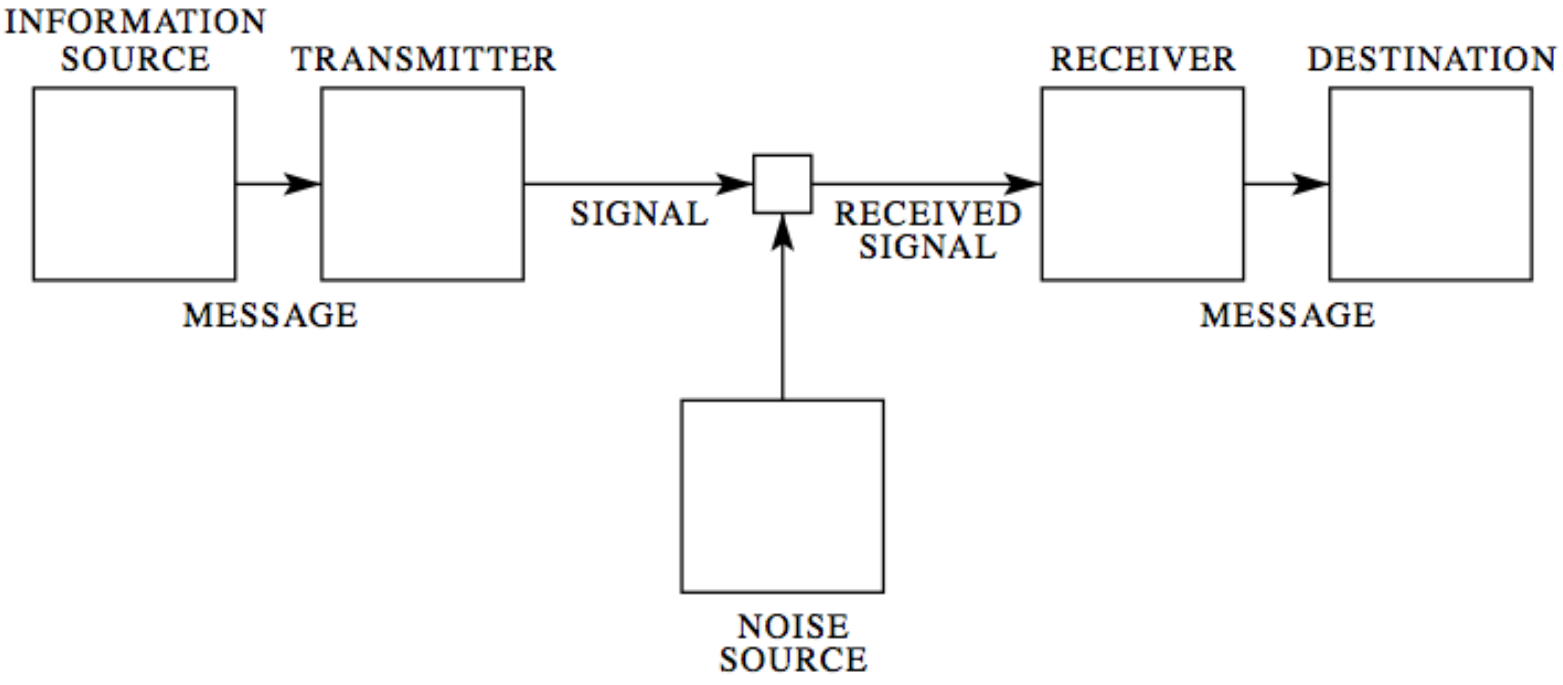


Fig. 1 — Schematic diagram of a general communication system.

Hlavní úlohy teorie informace

- komprese - efektivní reprezentace dat
- komunikace - spolehlivý přenos dat
- kryptografie - zabezpečení dat před nežádoucím přístupem

Rozlišujeme 2 hlavní třídy kódování:

1. zdrojové - při kompresi informačního zdroje odstraňujeme redundantní informaci
2. kanálové - větší redundance informace naopak zaručuje menší chybovost při přenosu

Co je informační zdroj?

Informační zdroj je pravděpodobnostní model zařízení, které produkuje zprávy složené ze znaků konečné abecedy Λ. Podle délky zpráv rozeznáváme 3 základní typy informačních zdrojů:

1. **náhodná veličina** X s výběrovým prostorem Λ
2. **náhodný vektor** $(X_1, \ldots, X_n)$ s výběrovým prostorem Λ^n
3. **náhodný proces** $(X_n)_{n \in \mathbb{N}}$ s výběrovým prostorem $\Lambda^{\mathbb{N}}$

Tyto informační zdroje generují zprávy dlouhé

- 1 znak
- n znaků
- (teoreticky) nekonečný počet znaků

Dva významy slova “bit”

1. symbol z množiny $\{0, 1\}$
2. jednotka informace při dvojkovém základu logaritmu

- podle Shannona zavedl termín “bit” známý statistik John W. Tukey již v roce 1947
- díky zvolené jednotce informace používáme dvojkový logaritmus a značení

$$\log := \log_2$$

Příklady informačních zdrojů

Příklad

$\Lambda = \{\mathtt{A}, \ldots, \mathtt{Z}, \_\}$.

Pravděpodobnosti $p_X(x)$ jednotlivých písmen $x \in \Lambda$ můžeme odhadnout např. pomocí dostatečně velkého souboru textů nad anglickou abecedou Λ. Viz tato studie.

Tento model však umožňuje popsat pouze informační zdroje, v němž jsou výskyty jednotlivých znaků $x_1, \ldots, x_n \in \Lambda$ ve slově $x_1 \ldots x_n$ **nezávislé**:

$$p_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_X(x_i).$$

Příklady informačních zdrojů (pokr.)

Příklad

$\Lambda = \{\mathtt{A}, \ldots, \mathtt{Z}, \_\}$.

Markovův řetězec $(X_n)_{n\in\mathbb{N}}$ popisuje slova nad abecedou Λ, kde pravděpodobnost dvojic písmen odpovídá typickému anglickému textu. Viz tato studie. Z toho lze odvodit pravděpodobnosti přechodu, např. platí

$$p_{\mathtt{TH}} > p_{\mathtt{TO}}, \quad p_{\mathtt{IN}} > p_{\mathtt{IT}}.$$

Shannon uvádí tento příklad realizace:

```
ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D
ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE
SEACE CTISBE
```

O čem to bude: komprese

Do telefonu si chceme uložit zkratky v binární abecedě $\{0, 1\}$ pro každé z 8 čísel přátel, kterým voláme s těmito pravděpodobnostmi:

$$\left(2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-6}, 2^{-6}, 2^{-6}, 2^{-6}\right)$$

Řešení

- je možno použít kód mající 3 bity pro každého
- lepší je však využít uvedené pravděpodobnosti ke konstrukci kódu kratší střední délky

Komprese: řešení

Řešení

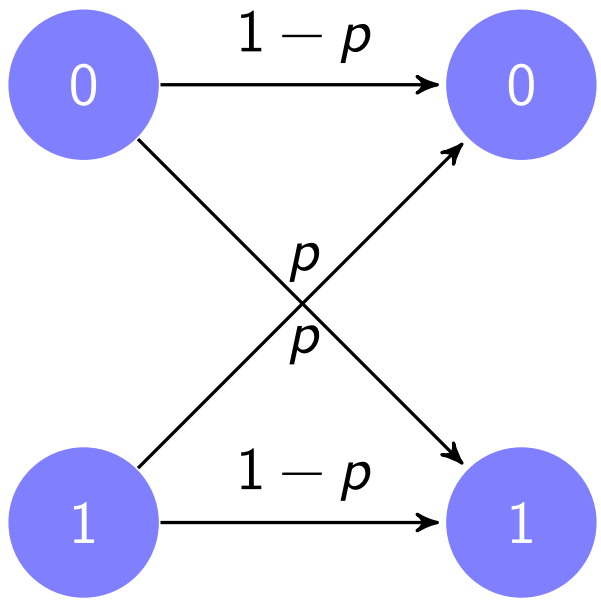
- střední délka kódu s 3-bitovými délkami kódových slov je 3
- uvažujme následující kód s délkami slov $\ell_i = -\log p_i$

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-6}	2^{-6}	2^{-6}
0	10	110	1110	111100	111110	111101	111111

- střední délka tohoto kódu je 2
- později uvidíme, že kód kratší střední délky nelze nalézt!

O čem to bude: komunikace

Chceme přenést bitové slovo informačním kanálem, v němž dochází vlivem šumu k záměně 1 bitu s pravděpodobností $p < 1/2$. Lze vhodným kódováním docílit menší pravděpodobnosti chyby λ_x záměny bitu x?



Komunikace: možné řešení

1. Kódování vstupu: každý bit x zopakuj $(2n+1)$-krát
2. Dekódování výstupu: podle většiny výstupních bitů $y_1 \dots y_{2n+1}$

To umožňuje opravit nejvýše n chyb, ovšem značně neefektivně:

- počet chyb E má rozdělení $\mathrm{Bi}(2n+1, p)$, tedy

$$\lambda_x = P[E > n] = \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} p^i (1-p)^{2n+1-i}$$

- při $n = 1$ to dává chybu $\lambda_x = 3p^2 - 2p^3 < p$
- když $n \to \infty$, potom $\lambda_x \to 0$
- ale $R \to 0$, kde $R := \frac{1}{2n+1}$ měří rychlost komunikace!

Aplikace teorie informace

Kompresní algoritmy

- Huffmanovo kódování, aritmetické kódování, LZW
- JPEG, MP3, TIFF

Kódy pro detekci a opravu chyb

- Reed-Solomonovy kódy (CD, DVD, DVB)
- turbo kódy (3G)

Z pohledu studenta FEL

Pochopím a naučím se používat

1. informační míry, které jsou důležité i mimo teorii informace (AI, strojové učení)
2. Huffmanovo kódování, které provádí optimální kompresi bezpaměťových zdrojů dat
3. model diskrétního kom. kanálu, který je základem pro moderní teorii kódování

Více v předmětu Teorie informace a kódování v LS 2014!

Část II

Charakteristiky informace

- Entropie
- Vzájemná informace
- Rychlost entropie

Hartleyho míra informace

Kolika bity vyjádříme n znaků?

Definice (Hartley, 1928)

Hartleyho míra informace I je funkce

$$I(n) = \log n, \quad n \in \mathbb{N}.$$

- platí-li $|\Lambda| = 2^k$ pro nějaké $k \geqslant 1$, potom $I(2^k) = k$ udává počet bitů, kterými lze zakódovat znak z Λ
- pokud neexistuje $k \geqslant 1$ takové, že $|\Lambda| = 2^k$, potom musíme kódovat znaky z Λ pomocí $\lceil I(|\Lambda|) \rceil$ bitů

Axiomy Hartleyho míry

Věta (Rényi)

Hartleyho míra informace je jediná funkce $I : \mathbb{N} \to \mathbb{R}$ následujících vlastností:

1. $I(m \cdot n) = I(m) + I(n)$, pro každé $m, n \in \mathbb{N}$
2. I je rostoucí
3. $I(2) = 1$

Intepretace

- k vyjádření prvku z $2^{k+\ell}$-prvkové množiny stačí zřetězit původní bitová slova délek k a ℓ
- počet bitů roste s počtem kódovaných znaků
- informaci měříme v bitech

Entropie: motivace

Pro popis každého z n výsledků potřebujeme informaci $I(n)$.

- $\Lambda = \{1, 2\}$, $p_1 = p_2 = \frac{1}{2}$. Pak $I(2) = \log 2 = \log \frac{1}{p_i} = 1$

Ovšem výsledky $i \in \Lambda$ mohou mít různou pravděpodobnost p_i!

- každý výsledek $i \in \Lambda$ přinese informaci $\log \frac{1}{p_i}$ bitů
- **průměrnou informaci**, kterou nám přinese znalost nějakého výsledku náhodného pokusu, tedy dostaneme jako

$$\sum_{i \in \Lambda} p_i \log \frac{1}{p_i} = -\sum_{i \in \Lambda} p_i \log p_i$$

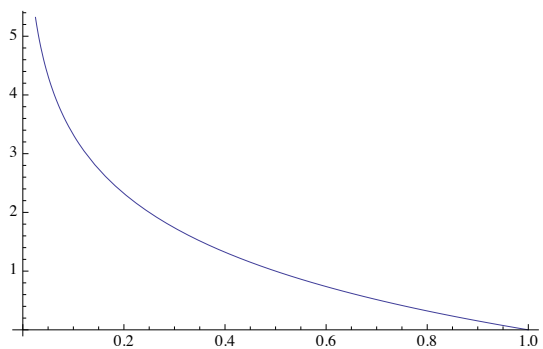
Entropie

Definice (Shannon)

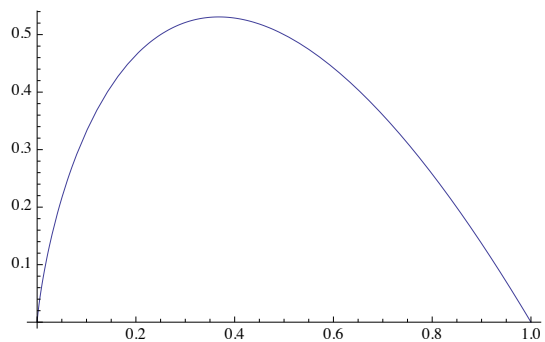
Nechť X je náhodná veličina s hodnotami v Λ a pravděpodobnostní funkcí p_X. Entropie náhodné veličiny X je

$$H(X) = -\sum_{x \in \Lambda} p_X(x) \log p_X(x),$$

kde užíváme konvenci $0 \log 0 = 0$.



Funkce $-\log p$ pro $p \in \langle 0, 1 \rangle$



Funkce $-p \log p$ pro $p \in \langle 0, 1 \rangle$

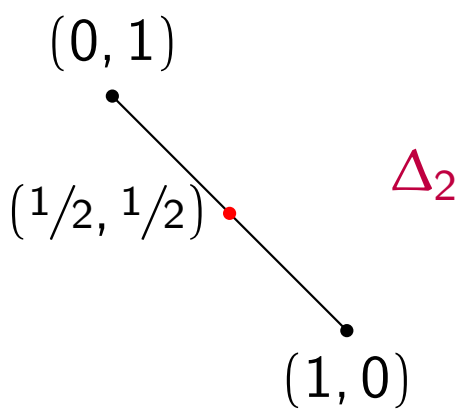
Entropie jako funkce

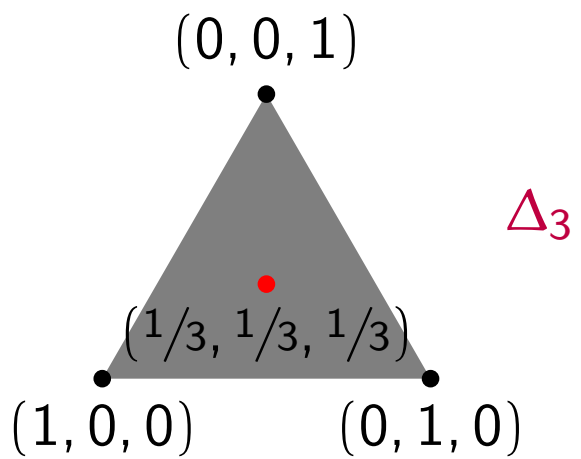
Nechť $|\Lambda| = n$. Entropii lze chápat jako reálnou funkci

$$H : \Delta_n \to \mathbb{R},$$

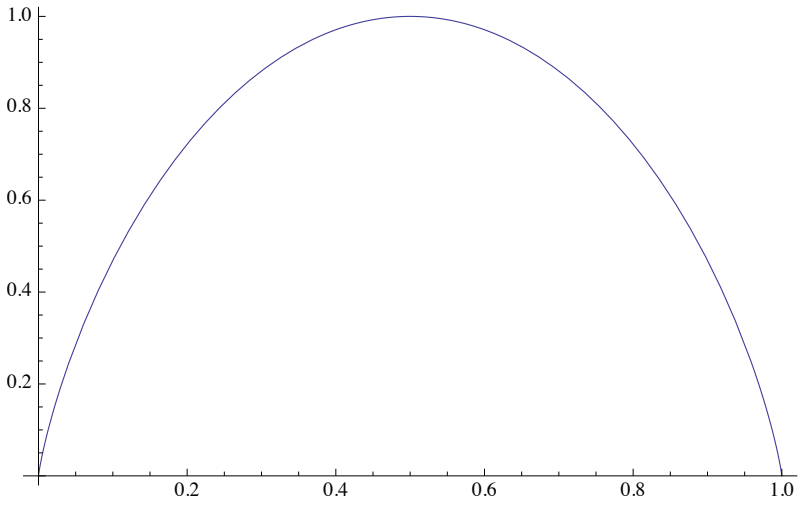
kde Δ_n je pravděpodobnostní n-simplex, což je množina všech pravděpodobnostních funkcí na Λ:

$$\Delta_n = \left\{ p \in \mathbb{R}^n \;\middle|\; p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

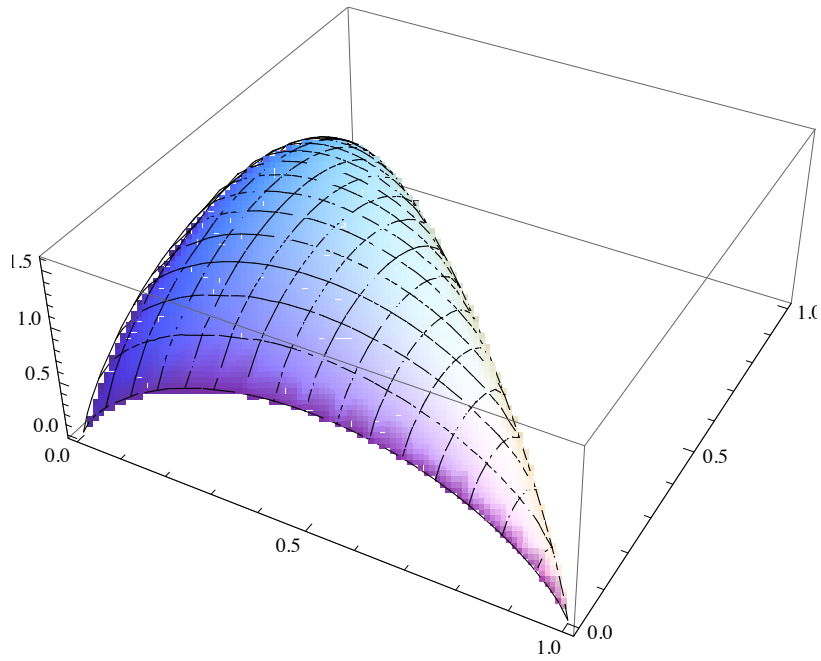




Entropie jako funkce (pokr.)



Entropie na Δ_2



Entropie na Δ_3

Axiomy entropie

Věta

Entropie $H(p_1, \ldots, p_n)$ je jediná funkce $\Delta_n \to \langle 0, \infty)$ následujících vlastností:

1. H nezávisí na pořadí argumentů $p_1, \ldots, p_n$
2. Funkce $(p, 1-p) \in \Delta_2 \mapsto H(p, 1-p)$ je spojitá
3. $H(1/2, 1/2) = 1$
4. $H(p_1, \ldots, p_n) =$
   $H(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2) H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$

První tři axiomy mají velmi přirozenou interpretaci. A co čtvrtý?

Interpretace 4. axiomu

Příklad

Úkolem je uhodnout soupeřem náhodně zvolené číslo z množiny $\{1, \ldots, 5\}$. Lze použít např. jednu z těchto strategií:

- náhodně vybereme číslo z množiny $\{1, \ldots, 5\}$, pokus je tak popsán rovnoměrným rozdělením $p_i = \frac{1}{5}, i = 1, \ldots, 5$
- nejprve volíme prvek množiny $\big\{\{1, 2\}, 3, 4, 5\big\}$ podle rozdělení $\left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$, potom případně náhodně volíme číslo z $\{1, 2\}$

Dostaneme vždy to samé:

- $H\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right) = I(5) = \log 5$
- $H\left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right) + \frac{2}{5} H\left(\frac{1}{2}, \frac{1}{2}\right) = \log 5$

Entropie o základu d

Informaci lze měřit i v jiných jednotkách než bitech ($d = 2$).

Věta

Nechť $d > 0$ a $H_d(X) = -\sum_{x \in \Lambda} p_X(x) \log_d p_X(x)$. Pak

$$H_d(X) = \log_d 2 \cdot H(X).$$

Důkaz.

Pro každé $t > 0$ platí rovnost $t = 2^{\log t}$. Jejím zlogaritmováním získáme vztah

$$\log_d t = \log_d 2 \cdot \log t,$$

ze kterého tvrzení plyne.

Minimální entropie

Výsledek pokusu v **Diracově rozdělení** nepřinese informaci.

Věta

Platí $H(X) \geqslant 0$. Rovnost $H(X) = 0$ nastane právě tehdy, když je rozdělení X Diracovo.

Důkaz.

- nerovnost plyne přímo z definice entropie
- pokud má X Diracovo rozdělení, potom $H(X) = \log 1 = 0$
- obráceně, z existence $x \in \Lambda$ s vlastností $1 > p_X(x) > 0$ plyne $-p_X(x) \log p_X(x) > 0$, a tudíž $H(X) > 0$

Maximální entropie

Výsledek popsaný **rovnoměrným rozdělením** přináší maximální možnou informaci.

Věta

Platí $H(X) \leqslant \log |\Lambda|$. Rovnost $H(X) = \log |\Lambda|$ nastane právě tehdy, když má X rovnoměrné rozdělení.

Důkaz.

Využijeme této nerovnosti platné pro každé $t > 0$:

$$\log t \leqslant (t-1) \log e,$$

kde rovnost nastává právě tehdy, když $t = 1$.

Sdružená a podmíněná entropie

Definice

Nechť (X, Y) je náhodný vektor s hodnotami v $\Lambda \times \Omega$. Sdružená entropie (X, Y) je

$$H(X, Y) = -\sum_{(x,y)\in\Lambda\times\Omega} p_{XY}(x, y) \log p_{XY}(x, y).$$

Pro dané $x \in \Lambda$, $p_X(x) > 0$ je podmíněná entropie dána jako

$$H(Y \mid X = x) = -\sum_{y\in\Omega} p_{Y|X}(y \mid x) \log p_{Y|X}(y \mid x)$$

a střední podmíněná entropie je $H(Y \mid X) = \sum_{x\in\Lambda} p_X(x) H(Y \mid X = x)$,
kde $H(Y \mid X = x)$ definujeme libovolně, pokud $p_X(x) = 0$.

Řetězcové pravidlo

Věta

Platí $H(X, Y) = H(X) + H(Y \mid X)$.

Důkaz.

Jelikož $p_{XY}(x, y) = p_X(x) p_{Y|X}(y|x)$,dostaneme

$$\begin{aligned} H(X, Y) &= -\sum_{(x,y)\in\Lambda\times\Omega} p_{XY}(x, y) \log p_X(x) p_{Y|X}(y \mid x) \\ &= -\sum_{(x,y)\in\Lambda\times\Omega} \big(p_{XY}(x, y) \log p_X(x) + p_{XY}(x, y) \log p_{Y|X}(y \mid x)\big) \\ &= -\sum_{x\in\Lambda} p_X(x) \log p_X(x) + \sum_{x\in\Lambda} H(Y \mid X = x) p_X(x). \end{aligned}$$

Interpretace řetězcového pravidla

Nechť X a Y popisují souřadnice pokladu na čtvercové síti.

- $H(X, Y)$ udává **průměrnou informaci** o pozici pokladu
- taková informace se k nám ovšem může dostat ve dvou kolech:

1. dozvíme se X-ovou souřadnici,
2. po odhalení X se dozvíme Y-ovou souřadnici.

Vztah je symetrický,

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

a lze ho snadno zobecnit pro více veličin:

$$H(X, Y, Z) = H(X) + H(Y|X) + H(Z|X, Y)$$

Entropie nezávislých veličin

Věta

Jsou-li X, Y nezávislé, pak

$$H(Y|X) = H(Y) \quad \text{a} \quad H(X,Y) = H(X) + H(Y).$$

Důkaz.

Důsledek řetězcového pravidla, vztahu

$$p_{Y|X}(y|x) = p_Y(y)$$

pro nezávislé veličiny X, Y a $H(Y|X = x) = H(Y)$.

- Entropie
- Vzájemná informace
- Rychlost entropie

Vzájemná informace

Jak závisí obdržená zpráva Y na zaslané zprávě X?

Definice

Nechť (X, Y) je náhodný vektor s hodnotami v $\Lambda \times \Omega$. Vzájemná informace $I(X; Y)$ je definována jako

$$I(X; Y) = \sum_{(x,y)\in\Lambda\times\Omega} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x) \cdot p_Y(y)}.$$

Interpretace

$I(X; Y)$ měří odlišnost sdruženého rozdělení náhodného vektoru (X, Y) od součinového rozdělení jeho marginálů, kterým by se vektor (X, Y) řídil, pokud by X a Y byly nezávislé.

$I(X; Y)$ je míra zachování informace

Věta

Platí $I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$.

Důkaz.

Protože $p_{XY} = p_{X|Y} p_Y$, dostaneme

$$\begin{aligned} I(X; Y) &= \sum_{(x,y) \in \Lambda \times \Omega} p_{XY}(x, y) \log \frac{p_{X|Y}(x \mid y)}{p_X(x)} \\ &= \sum_{(x,y) \in \Lambda \times \Omega} \big(-p_{XY}(x, y) \log p_X(x) + p_{XY}(x, y) \log p_{X|Y}(x \mid y) \big) \\ &= -\sum_{x \in \Lambda} p_X(x) \log p_X(x) - \sum_{y \in \Omega} H(X \mid Y = y) p_Y(y). \end{aligned}$$

Vlastnosti vzájemné informace

Věta

1. $I(X;Y) = I(Y;X)$
2. $0 \leqslant I(X;Y) \leqslant H(X)$
3. $I(X;Y) = 0 \Leftrightarrow X$ a Y jsou nezávislé
4. $I(X;Y) = H(X) \Leftrightarrow$ existuje $\delta : \Omega \to \Lambda$ splňující $p_{X|Y}(\delta(y) \mid y) = 1$ pro každé $y \in \Omega, p_Y(y) > 0$

Důkaz.

1. a 4. vlastnost plynou ihned z definic. Pro 2. a 3. využijeme opět nerovnosti platné pro každé $t > 0$:

$$\log t \leqslant (t-1) \log e,$$

kde rovnost nastává právě tehdy, když $t = 1$.

Příklad: bitová inverze

Příklad (maximum vzájemné informace)

$X \sim \mathrm{Bi}(1, 1/2)$
$Y = X \oplus 1$

Z toho plyne

$$p_{Y|X}(1|0) = p_{Y|X}(0|1) = 1$$

a tedy

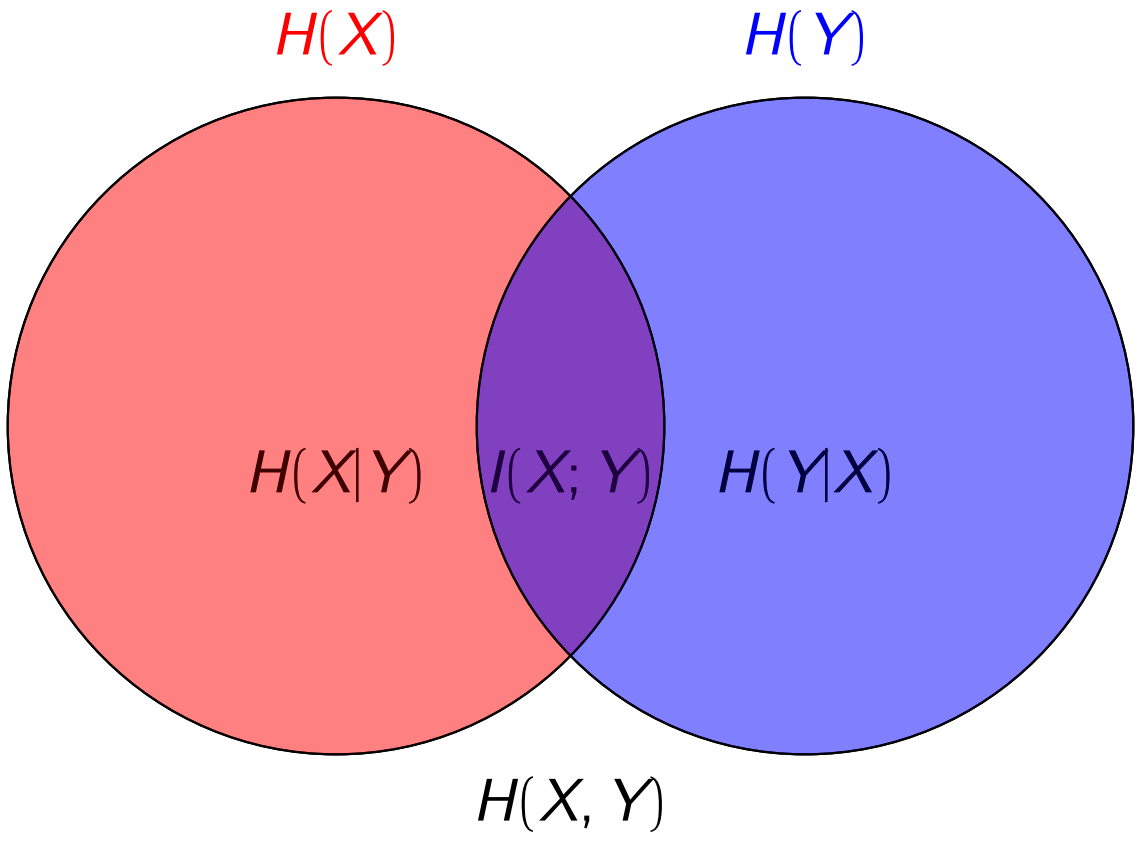
$$p_{X|Y}(1|0) = p_{X|Y}(0|1) = 1.$$

Stačí definovat

$$\delta(y) := y \oplus 1$$

a vidíme, že $I(X;Y) = H(X) = \log 2 = 1$.

Schéma



Význam informačních identit lze dobře ilustrovat na modelu informačního kanálu.

- Entropie
- Vzájemná informace
- Rychlost entropie

Motivace

Jak popíšeme informaci obsaženou ve zdroji $(X_n)_{n\in\mathbb{N}}$?

- zdroj $(X_1, \ldots, X_n)$ má sdruženou entropii $H(X_1, \ldots, X_n)$
- ovšem pro $n \to \infty$ není entropie zdroje $(X_1, \ldots, X_n)$ shora omezena, neboť pro $|\Lambda| \geqslant 2$ platí

$$H(X_1, \ldots, X_n) \leqslant \log |\Lambda|^n \to \infty$$

Řešení

Hledejme entropii na jeden znak zprávy!

Bezpaměťový zdroj

Definice

Bezpaměťový zdroj je informační zdroj $(X_n)_{n\in\mathbb{N}}$ nad konečnou abecedou Λ, kde veličiny $X_1, X_2, \ldots$ jsou **nezávislé** a **stejně rozdělené**.

Příklad: náhodné generování bitů

X_n je náhodný bit a $p_{X_n}(1) = p \in \langle 0, 1\rangle$, pro každé $n \in \mathbb{N}$

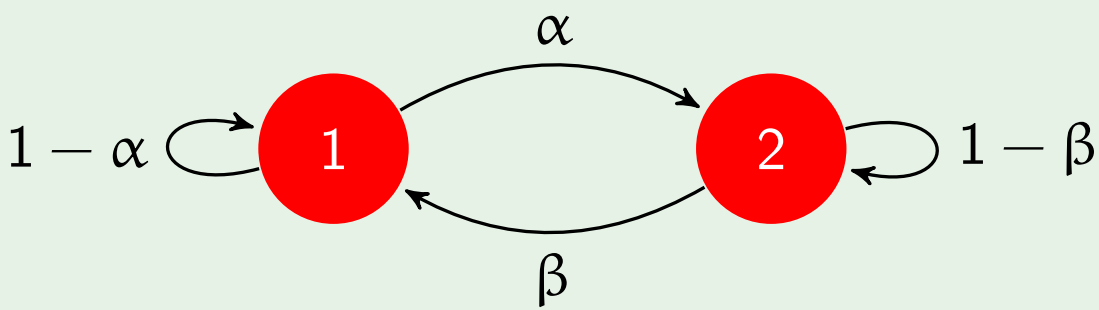
Příklad: náhodné generování textu

X_n je písmeno z abecedy $\Lambda = \{\mathtt{A}, \ldots, \mathtt{Z}, \_\}$, pravděpodobnosti jednotlivých písmen jsou stejné pro každé $n \in \mathbb{N}$

Markovův řetězec

Markovův řetězec $(X_n)_{n\in\mathbb{N}}$

stavový prostor $\Lambda = \{1, 2\}$
$\alpha, \beta \in \langle 0, 1\rangle$



Matice přechodu je $\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$ a počáteční rozdělení $\mathbf{p}(0)$ udává pravděpodobnosti stavů 1 a 2 na začátku.

Rychlost entropie

Definice

Rychlost entropie zdroje $(X_n)_{n\in\mathbb{N}}$ je

$$H((X_n)_{n\in\mathbb{N}}) = \lim_{n\to\infty} \frac{H(X_1 \ldots, X_n)}{n},$$

pokud tato limita existuje.

Příklad

Je-li zdroj zpráv $X_1, X_2, \ldots$ bezpaměťový, potom

$$H((X_n)_{n\in\mathbb{N}}) = H(X_1).$$

Mezní podmíněná entropie

Definice

Mezní podmíněná entropie $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ zdroje $(X_n)_{n\in\mathbb{N}}$ je limita posloupnosti

$$H(X_1),\ H(X_2|X_1),\ H(X_3|X_2,X_1),\ldots$$

pokud tato limita existuje.

Dostáváme tak 2 pojmy:

- $H((X_n)_{n\in\mathbb{N}})$ je entropie na znak vyslané zprávy
- $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ je entropie znaku při znalosti předchozích znaků

Oba pojmy splývají pro **stacionární** informační zdroje.

Stacionární informační zdroj

Definice

Zdroj $(X_n)_{n\in\mathbb{N}}$ je stacionární, pokud platí

$$P\big[X_1 = x_1, \ldots, X_n = x_n\big] = P\big[X_{1+\ell} = x_1, \ldots, X_{n+\ell} = x_n\big]$$

pro každé $n \in \mathbb{N}$, každý posun ℓ a každé $x_1, \ldots, x_n \in \Lambda$.

Příklad

Bezpaměťový zdroj

Zajímavější příklad

Markovův řetězec s maticí přechodu $\mathbf{P}$ a počátečním rozdělením $\mathbf{p}(0)$, které je stacionární: $\mathbf{p}(0) = \mathbf{p}(0)\mathbf{P}$.

Stacionarita a mezní podmíněná entropie

Věta

Je-li zdroj $(X_n)_{n\in\mathbb{N}}$ stacionární, potom je posloupnost $H(X_2|X_1),\ H(X_3|X_2,X_1),\ldots$ nerostoucí a její limita $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ existuje.

Důkaz.

Pro $n \geqslant 2$ platí

$$\begin{aligned} H(X_{n+1}|X_1,X_2,\ldots,X_n) &\leqslant H(X_{n+1}|X_2,\ldots,X_n) \\ &= H(X_n|X_1,\ldots,X_{n-1}) \end{aligned}$$

kde $\leqslant$ plyne z faktu, že podmiňování snižuje entropii a $=$ je důsledkem stacionarity. Dostaneme tak nerostoucí posloupnost nezáporných reálných čísel, jejíž limita $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ existuje.

Stacionarita a rychlost entropie

Věta

Je-li zdroj $(X_n)_{n\in\mathbb{N}}$ stacionární, potom rychlost entropie $H((X_n)_{n\in\mathbb{N}})$ existuje a platí

$$H((X_n)_{n\in\mathbb{N}}) = \widetilde{H}((X_n)_{n\in\mathbb{N}}).$$

Důkaz.

Podle **řetězcového pravidla** pro sdruženou entropii platí

$$\frac{H(X_1,\ldots,X_n)}{n} = \frac{\sum_{i=2}^{n} H(X_i|X_1,\ldots,X_{i-1}) + H(X_1)}{n}.$$

Vpravo je průměr posloupnosti, která konverguje k $\widetilde{H}((X_n)_{n\in\mathbb{N}})$. Tudíž musí konvergovat k $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ i posloupnost průměrů.

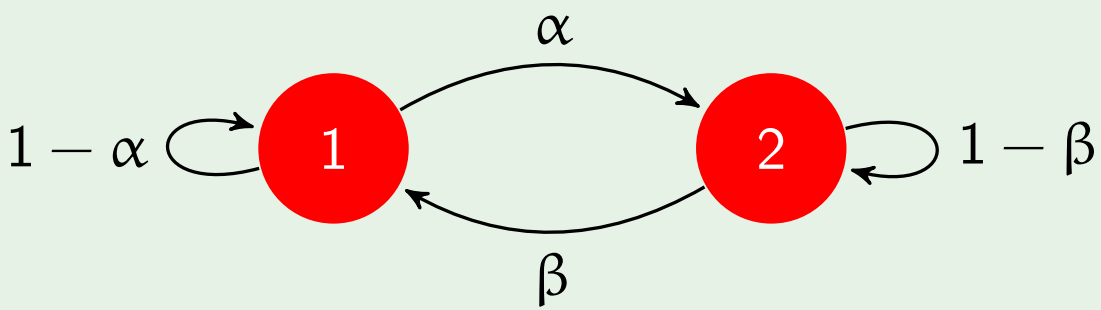
Markovský zdroj

Definice

Markovský zdroj informace je stacionární Markovův řetězec.

Příklad (pokr.)

Je-li $\alpha + \beta = 0$, pak je libovolné počáteční rozdělení $\mathbf{p}(0)$ stacionární. Předpokládejme $\alpha + \beta > 0$.



Existuje jediné stacionární rozdělení $\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ a tak klademe $\mathbf{p}(0) = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$.

Rychlost entropie markovského zdroje

Věta

Pokud je zdroj informace $(X_n)_{n\in\mathbb{N}}$ markovský, potom

$$H((X_n)_{n\in\mathbb{N}}) = H(X_2|X_1).$$

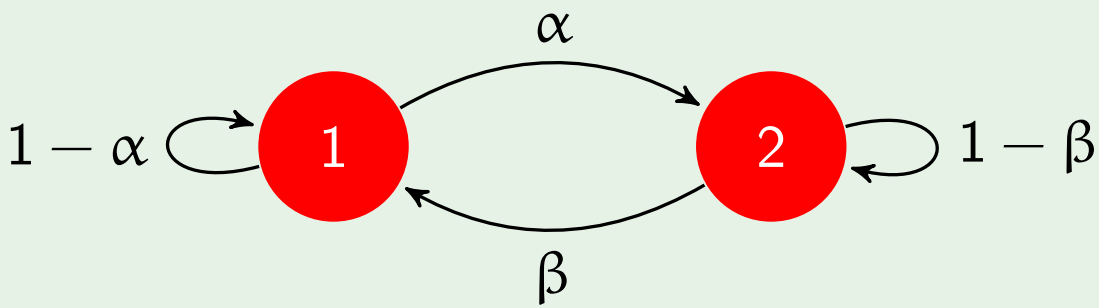
Důkaz.

$$\begin{aligned} H((X_n)_{n\in\mathbb{N}}) &= \widetilde{H}((X_n)_{n\in\mathbb{N}}) = \lim_{n\to\infty} H(X_{n+1}|X_n,\ldots,X_1) \\ &= \lim_{n\to\infty} H(X_{n+1}|X_n) = H(X_2|X_1). \end{aligned}$$

Příklad markovského zdroje

Příklad (pokr.)

Je-li $\alpha + \beta = 0$, pak $H((X_n)_{n\in\mathbb{N}}) = 0$. Nechť $\alpha + \beta > 0$.

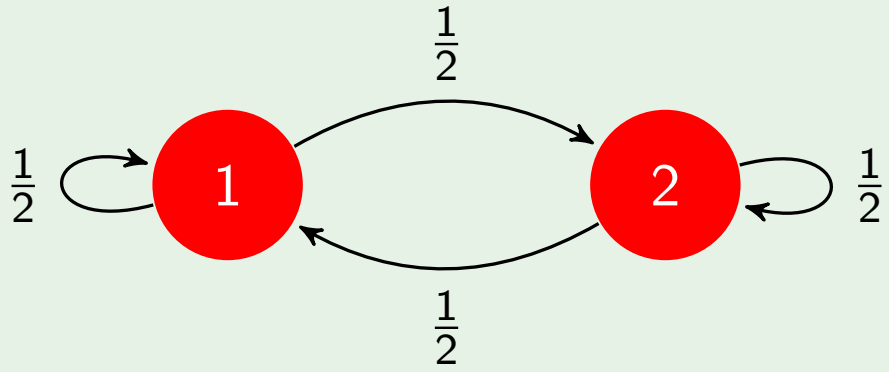


Počáteční rozdělení $\mathbf{p}(0) = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$. Rychlost entropie je

$$H((X_n)_{n\in\mathbb{N}}) = H(X_2|X_1) = \frac{\beta}{\alpha+\beta} H(\alpha, 1-\alpha) + \frac{\alpha}{\alpha+\beta} H(\beta, 1-\beta).$$

Příklad markovského zdroje: $\alpha = \beta = \frac{1}{2}$

Příklad: $H((X_n)_{n\in\mathbb{N}}) = 1$



Počáteční rozdělení $\mathbf{p}(0) = \left(\frac{1}{2}, \frac{1}{2}\right)$

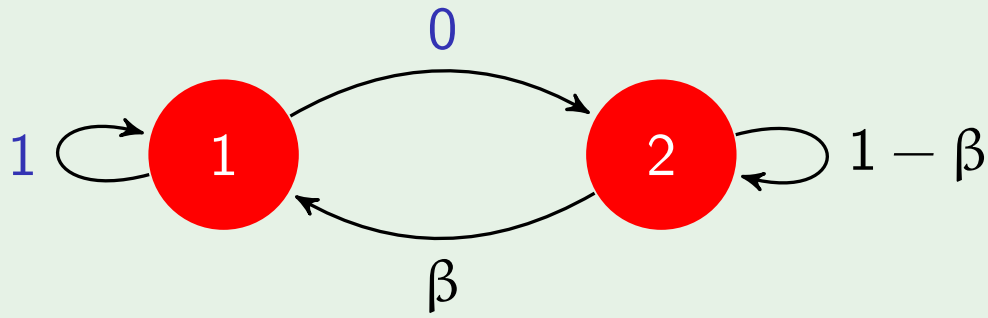
Jde o **bezpaměťový** zdroj informace a ve zprávě $x_1 x_2 \ldots$ má každý řetězec

$$x_k x_{k+1} \ldots x_{k+\ell}$$

stejnou pravděpodobnost $2^{-\ell-1}$.

Příklad markovského zdroje: $\alpha = 0,\ \beta > 0$

Příklad: $H((X_n)_{n\in\mathbb{N}}) = 0$



Počáteční rozdělení $\mathbf{p}(0) = (1, 0)$

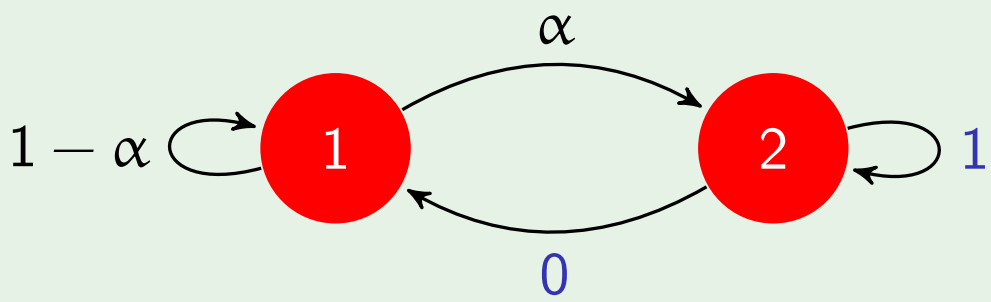
Jde o **deterministický** zdroj informace: zpráva

$$111\ldots$$

vznikne s pravděpodobností 1.

Příklad markovského zdroje: $\beta = 0,\ \alpha > 0$

Příklad: $H((X_n)_{n\in\mathbb{N}}) = 0$



Počáteční rozdělení $\mathbf{p}(0) = (0, 1)$

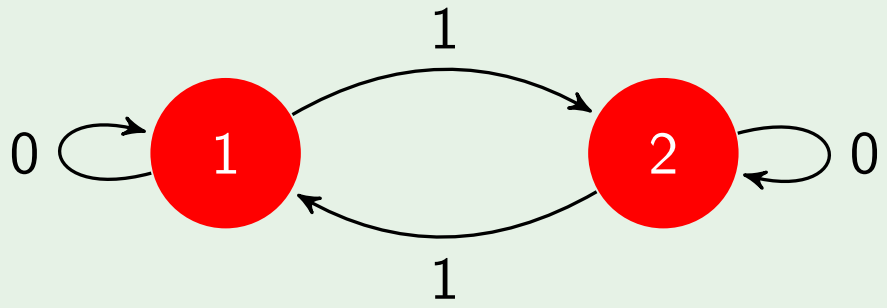
Jde o **deterministický** zdroj informace: zpráva

$$222\ldots$$

vznikne s pravděpodobností 1.

Příklad markovského zdroje: $\alpha = \beta = 1$

Příklad: $H((X_n)_{n\in\mathbb{N}}) = 0$



Počáteční rozdělení $\mathbf{p}(0) = \left(\frac{1}{2}, \frac{1}{2}\right)$, oba stavy mají periodu 2

Jde o **nedeterministický** zdroj informace, ale každá zpráva je jednoznačně určena prvním znakem: zprávy

$$1212\ldots \quad \text{a} \quad 2121\ldots$$

mají obě pravděpodobnost $1/2$.

Část III

Asymptotická rovnočetnost typických zpráv

- Prolog k bezeztrátové kompresi
- Typické zprávy

Motivace

Příklad

Bezpaměťový zdroj generuje 100 bitů s pravděpodobnostmi $p(0) = 0.995$, $p(1) = 0.005$. Máme zakódovat každou zprávu $x_1 \ldots x_{100}$, která obsahuje nejvýše tři bity 1.

- počet **typických zpráv** (nejvýše tři bity 1) je

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 166\,751$$

- tudíž stačí kódová slova délky $\lceil \log 166\,751 \rceil = 18$
- tak zakódujeme zprávy z množiny, jejíž pravděpodobnost je

$$\sum_{i=0}^{3} \binom{100}{i} 0.005^i \, 0.995^{100-i} = 0.99833$$

Motivace (pokr.)

Příklad (pokr.)

- tím ovšem dosáhneme jen ztrátové komprese
- **atypické zprávy** (více než tři 1) kódujeme pomocí původních

$$100 \text{ bitů}$$

- přidáním počátečního **příznakového** bitu odlišíme kódová slova pro typické a atypické zprávy
- **střední délka** kódového slova je tak

$$0.99833 \cdot 19 + 0.00167 \cdot 101 = 19.13694$$

Motivace (pokr.)

Jak poznáme **typické zprávy**?

- pravděpodobnost typické zprávy $(\underbrace{0,\ldots,0}_{100\times})$ je

$$p_{100} = 0.995^{100} = 0.60577$$

- povšimněme si, že existuje $\varepsilon > 0$ takové, že

$$p_{100} \approx 2^{-100\left(H(0.005,0.995)-\varepsilon\right)}$$

- pravděpodobnost každé typické zprávy je tedy přibližně

$$2^{-100H(0.005,0.995)}$$

Obecná úloha

Uvažujme tyto informační zdroje:

- X s pravděpodobnostní funkcí p na množině $\mathcal{X}$
- $(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, popsaný pomocí sdružené pravděpodobnostní funkce

$$p_n(\mathbf{x}^n) = \prod_{i=1}^{n} p(x_i), \quad \mathbf{x}^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$$

na množině $\mathcal{X}^n$

Interpretace

- zdroj $(X_1, \ldots, X_n)$ je **bezpaměťový**
- pravděpodobnost výskytu znaku $x \in \mathcal{X}$ je vždy $p(x)$

Obecná úloha (pokr.)

Pro dlouhé zprávy řešme tuto úlohu bezeztrátové komprese:

- hledáme **binární blokový kód** pro $(X_1, \ldots, X_n)$: kódujeme celé bloky zpráv $x_1 \ldots x_n$, nikoli jednotlivé znaky x_i!
- tento kód musí umožňovat jednoznačnou rekonstrukci původních zpráv z $\mathcal{X}^n$
- cílem je minimalizovat střední délku tohoto kódu

AEP (Asymptotic Equipartition Property)

- využití slabého zákona velkých čísel v teorii informace
- efektivní komprese zpráv z **typické množiny**

Typičnost

Definice

Nechť $\varepsilon > 0$. Množina ε-typických zpráv zdroje $(X_1, \ldots, X_n)$ je

$$A_\varepsilon^{(n)} = \left\{ \mathbf{x}^n \in \mathcal{X}^n \;\middle|\; 2^{-n(H(X)+\varepsilon)} \leqslant p_n(\mathbf{x}^n) \leqslant 2^{-n(H(X)-\varepsilon)} \right\}.$$

Ekvivalentně: zpráva $\mathbf{x}^n \in \mathcal{X}^n$ je ε-typická $\Leftrightarrow$

$$H(X) - \varepsilon \leqslant -\frac{1}{n} \log p_n(\mathbf{x}^n) \leqslant H(X) + \varepsilon.$$

Typické zprávy umožňují “dobrý” odhad entropie $H(X)$.

AEP

Věta

Nechť $\varepsilon > 0$.

1. Pro každé $n \in \mathbb{N}$ platí
$$|A_\varepsilon^{(n)}| \leqslant 2^{n(H(X)+\varepsilon)}.$$
2. Existuje n_ε takové, že
$$P(A_\varepsilon^{(n)}) > 1-\varepsilon, \quad \text{pro } n > n_\varepsilon.$$
3. $|A_\varepsilon^{(n)}| \geqslant (1-\varepsilon)2^{n(H(X)-\varepsilon)}$, pro $n > n_\varepsilon$.

Důkaz AEP

Důkaz.

První tvrzení:

$$1 = \sum_{\mathbf{x}^n \in \mathcal{X}^n} p_n(\mathbf{x}^n) \geqslant \sum_{\mathbf{x}^n \in A_\varepsilon^{(n)}} p_n(\mathbf{x}^n)$$
$$\geqslant \sum_{\mathbf{x}^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)} |A_\varepsilon^{(n)}|.$$

Druhé tvrzení: položme $Y_i = -\log p_{X_i}(x_i)$. Pak $E(Y_i) = H(X)$ a

$$P(A_\varepsilon^{(n)}) = P\big[|\bar{Y}_n - H(X)| \leqslant \varepsilon\big] = P\big[|\bar{Y}_n - E(Y_i)| \leqslant \varepsilon\big].$$

Nerovnost tak plyne přímo ze slabého ZVČ.

Důkaz AEP (pokr.)

Důkaz.

Třetí tvrzení: pro $n > n_\varepsilon$ platí

$$1 - \varepsilon < P(A_\varepsilon^{(n)}) = \sum_{\mathbf{x}^n \in A_\varepsilon^{(n)}} p_n(\mathbf{x}^n)$$
$$\leqslant \sum_{\mathbf{x}^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)} |A_\varepsilon^{(n)}|.$$

Využití AEP

1. Horní odhad velikosti $A_\varepsilon^{(n)}$
2. Množina typických zpráv má pravděpodobnost 1 pro $n \to \infty$
3. Dolní odhad velikosti $A_\varepsilon^{(n)}$ (závislý na $n > n_\varepsilon$)

Kódování pomocí AEP

- přednostní pozornost budeme věnovat typickým zprávám
- takový přístup ale bude stačit k výhodné kompresi všech zpráv!

Kódování pomocí AEP

1. každou zprávu z $A_\varepsilon^{(n)}$ lze zakódovat pomocí nejvýše $n(H(X)+\varepsilon)+1$ bitů, přidáním počátečního **příznakového bitu** 0 je horní mez $n(H(X)+\varepsilon)+2$ bitů
2. každou zprávu z $\mathcal{X}^n \setminus A_\varepsilon^{(n)}$ zakódujeme pomocí nejvýše $n \log |\mathcal{X}| + 2$ bitů

Vlastnosti

- kód umožňuje dékodování **ihned** po obdržení kódového slova
- dosáhneme **efektivní komprese**: každé kódové slovo je v průměru vyjádřeno pomocí $nH(X)$ bitů

Efektivita kódování pomocí AEP

Věta

Nechť $\ell(X^n)$ značí délku kódového slova pro zprávu X^n. Pro každé $\hat{\varepsilon} > 0$ platí

$$E\left(\frac{1}{n}\ell(X^n)\right) - H(X) < \hat{\varepsilon}, \quad \text{pro } n \to \infty.$$

- uvedená věta platí pro nezávislé a stejně rozdělené veličiny $X_1, X_2, \ldots$, ale existuje zobecnění pro jisté stacionární zdroje
- střední délku kódového slova na jeden znak lze tak libovolně přiblížit k entropii původního zdroje

Efektivita kódování pomocí AEP (pokr.)

Důkaz.

Díky 2.tvrzení AEP platí pro $n > n_0$ a $\varepsilon > 0$:

$$
\begin{aligned}
E\left(\tfrac{1}{n}\ell(X^n)\right) &= \tfrac{1}{n}\sum_{\mathbf{x}^n\in\mathcal{X}^n} p_n(\mathbf{x}^n)\ell(\mathbf{x}^n) \\
&= \sum_{\mathbf{x}^n\in A_\varepsilon^{(n)}} p_n(\mathbf{x}^n)\left(H(X)+\varepsilon+\tfrac{2}{n}\right) + \sum_{\mathbf{x}^n\in\mathcal{X}^n\setminus A_\varepsilon^{(n)}} p_n(\mathbf{x}^n)\left(\log|\mathcal{X}|+\tfrac{2}{n}\right) \\
&= \left(H(X)+\varepsilon+\tfrac{2}{n}\right)P(A_\varepsilon^{(n)}) + \left(\log|\mathcal{X}|+\tfrac{2}{n}\right)\left(1-P(A_\varepsilon^{(n)})\right) \\
&< H(X) + \underbrace{\varepsilon+\tfrac{2}{n}+\varepsilon\left(\log|\mathcal{X}|+\tfrac{2}{n}\right)}_{\hat{\varepsilon}}
\end{aligned}
$$

Ilustrace AEP

Příklad

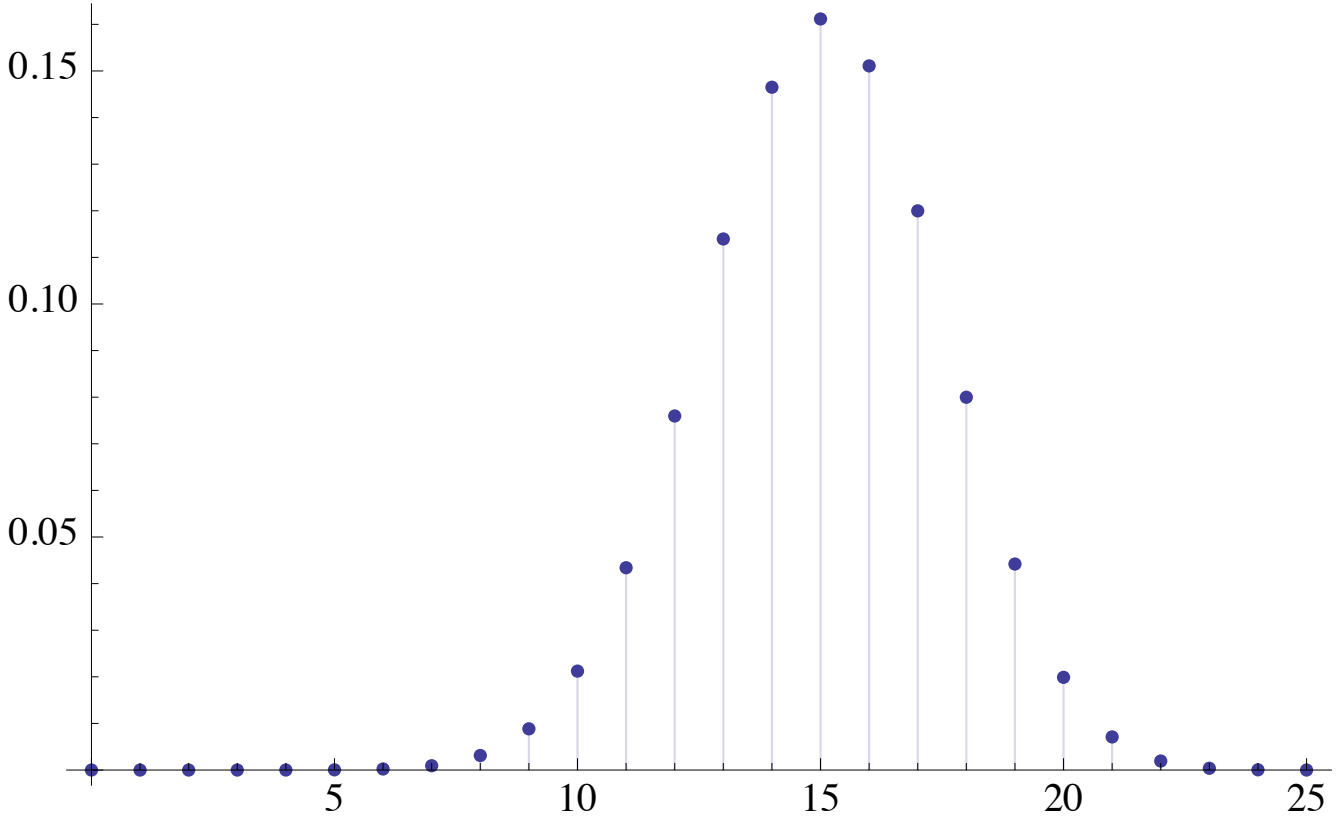
Bezpaměťový zdroj generuje $n = 25$ bitů s pravděpodobnostmi $p(0) = 0.4$, $p(1) = 0.6$. Zřejmě $H(X) \doteq 0.97$. Pro $\varepsilon = 0.1$ určíme

1. $A_\varepsilon^{(n)} = \left\{ \mathbf{x}^n \in \mathcal{X}^n \;\middle|\; 0.96 \leqslant -\frac{1}{n} \log p_n(\mathbf{x}^n) \leqslant 0.98 \right\}$
2. $P(A_\varepsilon^{(n)})$
3. $|A_\varepsilon^{(n)}|$

Stačí vyhodnotit počet jednotkových bitů ve zprávě délky 25 bitů.

Varování: nejpravděpodobnější zpráva (samé 1) zde není typická!

Ilustrace AEP (pokr.)



Pravděpodobnostní funkce Bi(25,0.6)

Ilustrace AEP (pokr.)

Příklad (pokr.)

1. $A_\varepsilon^{(n)} = \left\{ (x_1, \ldots, x_{25}) \in \{0,1\}^{25} \;\middle|\; 11 \leqslant \sum_{i=1}^{25} x_i \leqslant 19 \right\}$
2. $P(A_\varepsilon^{(n)}) = \sum_{i=11}^{19} \binom{25}{i} 0.6^i \, 0.4^{25-i} = 0.93625 > 1 - \varepsilon = 0.9$
3. $|A_\varepsilon^{(n)}| = \sum_{i=11}^{19} \binom{25}{i} = 26\,366\,510 < |\mathcal{X}^n| = 2^{25} = 33\,554\,432$

Odhad velikosti $A_\varepsilon^{(n)}$ pomocí AEP je

$$3\times10^6 \approx (1-\varepsilon)2^{n\left(H(0.6,0.4)-\varepsilon\right)} \leqslant |A_\varepsilon^{(n)}| \leqslant 2^{n\left(H(0.6,0.4)+\varepsilon\right)} \approx 1.15\times10^8$$

Část IV

Komprese zpráv a kódování

- Prolog k bezeztrátové kompresi
- Typické zprávy

Kódování

Definice

Nechť X je náhodná veličina s hodnotami v Λ. Nechť $\mathcal{D}$ je konečná abeceda a $\mathcal{D}^* = \bigcup_{n=1}^{\infty} \mathcal{D}^n$. Kód pro X je zobrazení

$$C : \Lambda \to \mathcal{D}^*.$$

Poznámka. Pro $\mathcal{D} = \{0, \ldots, d-1\}$, $d \geqslant 2$ hovoříme o d-znakovém kódu. Fyzikální principy přenosu a uchování dat vedou k využití zejména binárních kódů ($d = 2$).

Definice

Střední délka kódu C je $L(C) = \sum_{x \in \Lambda} p_X(x) \cdot \ell(C(x))$, kde $\ell(C(x))$ značí délku kódového slova $C(x)$.

Příklady kódů

Příklad 1

Nechť $p_X(i) = 1/3,\ i = 1, 2, 3$. Mějme tento binární kód:

$$C(1) = 0, C(2) = 10, C(3) = 11.$$

Zřejmě $L(C) = 5/3 = 1.\overline{6}$, přičemž $H(X) = \log 3 \doteq 1.58$.

Příklad 2

Nechť $p_X(i) = 2^{-i},\ i = 1, 2, 3$ a $p_X(4) = 2^{-3}$. Mějme tento binární kód:

$$C(1) = 0, C(2) = 10, C(3) = 110, C(4) = 111.$$

Zřejmě $L(C) = H(X) = 1.75$.

Třídy kódů

Definice

Kód C je

- nesingulární, pokud je $C : \Lambda \to \mathcal{D}^*$ prosté zobrazení.
- jednoznačně dekódovatelný, pokud je jeho rozšíření C^* nesingulární, kde $C^* : \Lambda^* \to \mathcal{D}^*$ je definováno pomocí

$$C^*(x_1 \dots x_n) = C(x_1) \dots C(x_n), \quad x_1 \dots x_n \in \Lambda^*.$$

- instantní, pokud žádné kódové slovo $C(x)$ není počátečním úsekem kódového slova $C(x')$ pro $x, x' \in \Lambda$, $x \neq x'$.

Třídy kódů (pokr.)

Věta (Vztahy mezi kódy)

1. Každý instantní kód je jednoznačně dekódovatelný.
2. Každý jednoznačně dekódovatelný kód je nesingulární.

Důkaz.

Druhé tvrzení plyne přímo z definice rozšíření C^*.

Nechť C není jednoznačně dekódovatelný. Potom C^* je singulární, tedy $C^*(x_1 \dots x_m) = C^*(y_1 \dots y_n)$ pro nějaké dvě zprávy $x_1 \dots x_m$, $y_1 \dots y_n \in \Lambda^*$, kde $x_1 \dots x_m \neq y_1 \dots y_n$. Kratší ze slov $C(x_1)$, $C(y_1)$ musí být tudíž prefixem delšího z nich.

Třídy kódů (pokr.)

x	singulární	nesingulární	jedn. dekódovatelný	instantní
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111

- pro **nesingulární kód** může být kódové slovo 010 dekódováno jako 2, 14 nebo 31
- kód ve 4. sloupci je **jednoznačně dekódovatelný**:
  1. pokud jsou první dva bity 00 nebo 10, lze je dekódovat
  2. pokud jsou první dva bity 11, potom
     - 2-1 je-li další bit 1, pak je první znak zprávy 3
     - 2-2 je-li délka řetězce nulových bitů za 11 lichá, pak je první znak zprávy 4
     - 2-3 je-li délka řetězce nulových bitů za 11 sudá, pak je první znak zprávy 3
  3. na další znaky kódového slova použijeme stejný postup

Instantní kódy

- **neinstantní jednoznačně dekódovatelný kód** obecně umožňuje dekódovaní až po přečtení celého rozšířeného kódového slova $C^*(x_1 \dots x_n)$
- **instantní kód** umožňuje dekódování ihned po obdržení kódového slova

Příklad

$\Lambda = \{1, 2, 3, 4\}$, $C(1) = 0$, $C(2) = 10$, $C(3) = 110$, $C(4) = 111$
Potom rozšířené kódové slovo

$$01011111010$$

kóduje zprávu 12432, tedy $C(12432) = 01011111010$.

Jak konstruovat kódy?

Úloha. Pro náhodnou veličinu X hledáme kód C, jehož střední délka $L(C)$ je minimální

- ukážeme, že při hledání minima se lze omezit na množinu **instantních kódů**
- při konstrukci instantního kódu nelze přiřazovat krátká kódová slova všem znakům zdrojové abecedy Λ, neboť to by vedlo k porušení instantnosti
- existuje **dolní mez** pro střední délku libovolného kódu, kterou nelze překročit
- **Kraftova nerovnost** nám umožní nalézt délky slov instantního kódu

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování
- Zdroje s pamětí

Kraftova nerovnost

Věta (Kraft, 1949)

Délky slov $\ell_1, \ldots, \ell_m$ libovolného **instantního** d-znakového kódu splňují nerovnost

$$\sum_{i=1}^{m} d^{-\ell_i} \leqslant 1.$$

Obráceně, splňují-li $\ell_1, \ldots, \ell_m \in \mathbb{N}$ tuto nerovnost, potom existuje **instantní** d-znakový kód s délkami slov $\ell_1, \ldots, \ell_m$.

Příklad

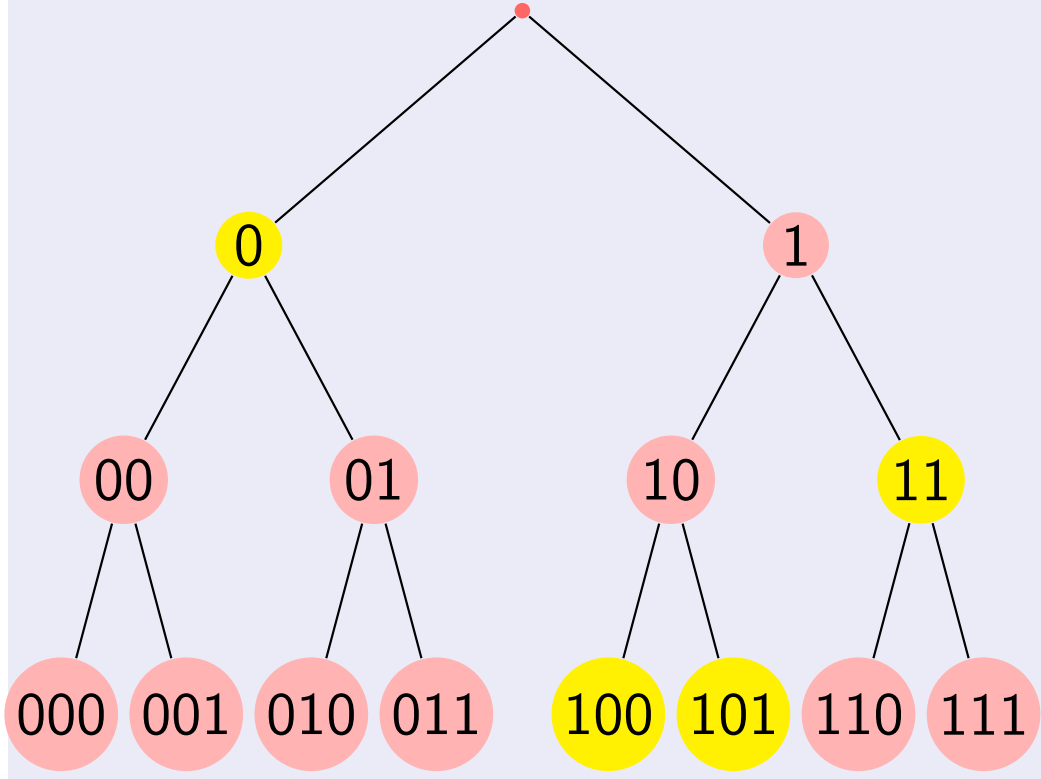
Kraftova nerovnost je splněna pro exponenty $\ell_1, \ldots, \ell_m$ každé dyadické pravděpodobnostní funkce p, tedy p splňující $p(x_i) = 2^{-\ell_i}$, $i = 1, \ldots, m$. Např. pro $m = 3$ jsou takové koeficienty pouze $1, 2, 2$.

Důkaz Kraftovy nerovnosti

Větu dokážeme pro binární kódy ($d = 2$).

Důkaz.

K instantnímu kódu zkonstruujeme kódovací strom:



V kódovacím stromě jsou kódová slova vyznačena žlutě. Jelikož je kód **instantní**, žádné kódové slovo není následníkem jiného kódového slova.

Důkaz Kraftovy nerovnosti (pokr.)

Pokračování důkazu.

Nechť ℓ je délka nejdelšího kódového slova. Potom:

- každé kódové slovo v úrovni ℓ_i má právě $2^{\ell-\ell_i}$ následníků v úrovni ℓ
- tyto množiny následníků jsou vzájemně disjunktní pro všechna kódová slova díky **instantnosti** kódu
- počet uzlů v úrovni ℓ je 2^ℓ, tedy dostáváme

$$\sum_{i=1}^{m} 2^{\ell-\ell_i} \leqslant 2^\ell, \qquad \text{což dává} \qquad \sum_{i=1}^{m} 2^{-\ell_i} \leqslant 1$$

Důkaz Kraftovy nerovnosti (pokr.)

Pokračování důkazu.

Obráceně, nechť platí Kraftova nerovnost a $\ell_1 \leqslant \cdots \leqslant \ell_m$. Mějme **binární strom** hloubky ℓ_m:

- v úrovni ℓ_1 vyznačme libovolný uzel a vyjměme jeho následníky
- nechť byl vyznačen uzel v úrovni ℓ_i $(1 \leqslant i \leqslant m-1)$ a jeho následníci vyjmuti
- nalezneme uzel v úrovni ℓ_{i+1}?Je zde

$$2^{\ell_{i+1}} - 2^{\ell_{i+1}-\ell_1} - \ldots - 2^{\ell_{i+1}-\ell_i} \quad \text{uzlů}$$

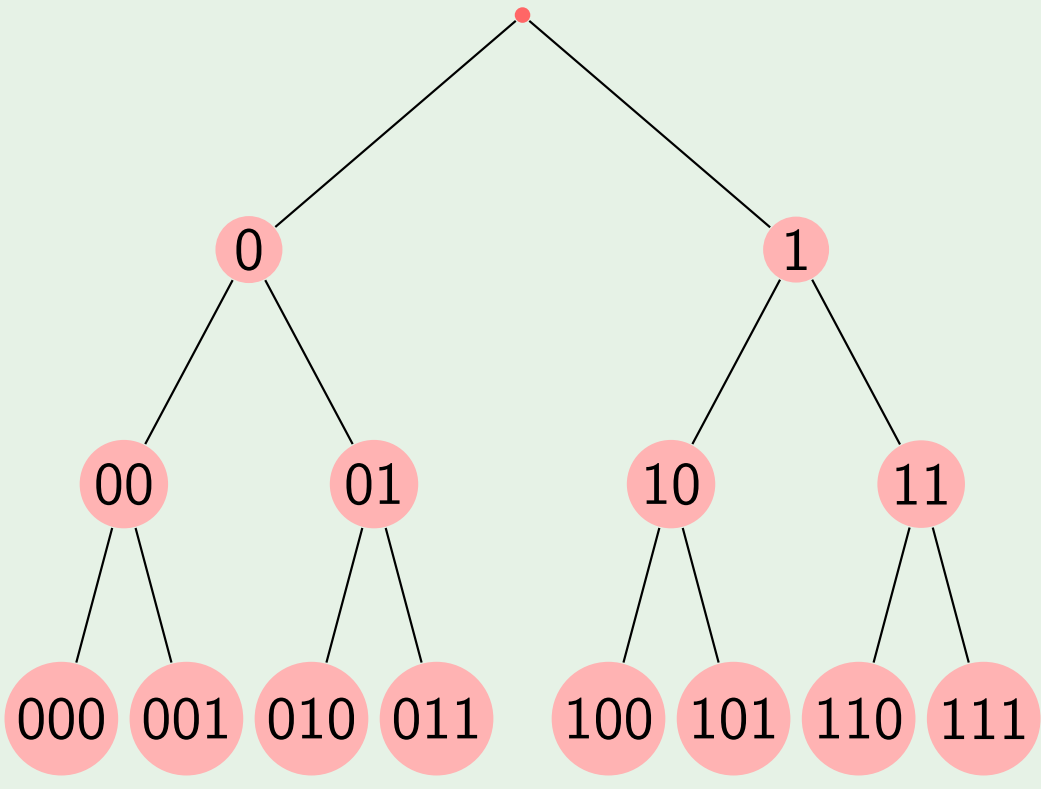
- tento počet je $\geqslant 1$ díky Kraftově nerovnosti (násobme $2^{\ell_{i+1}}$):

$$\sum_{j=1}^{i+1} 2^{-\ell_j} \leqslant 1$$

Ilustrace Kraftovy nerovnosti

Příklad

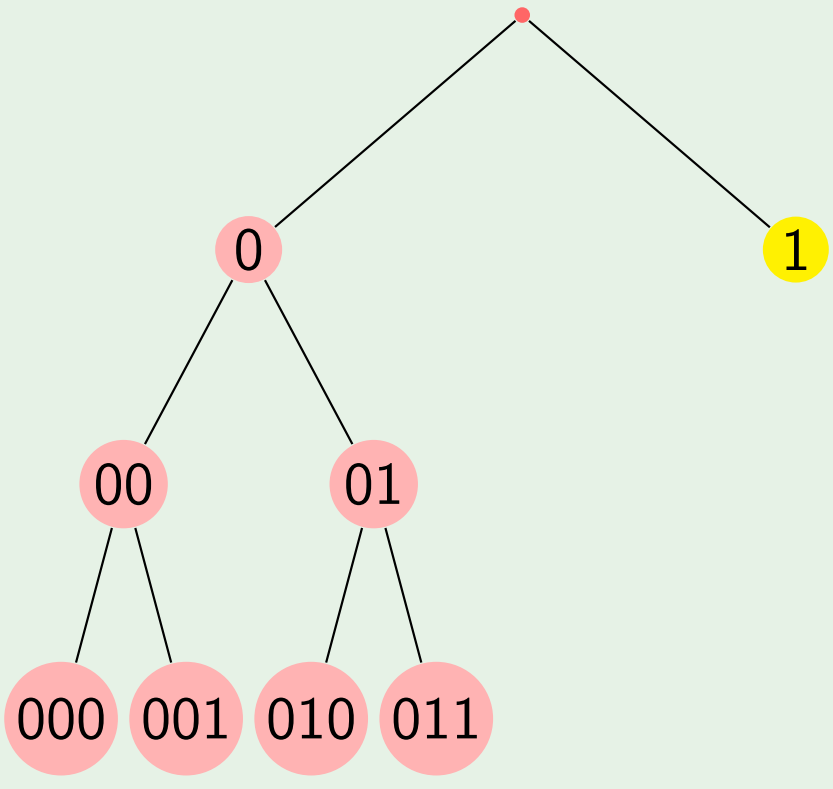
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

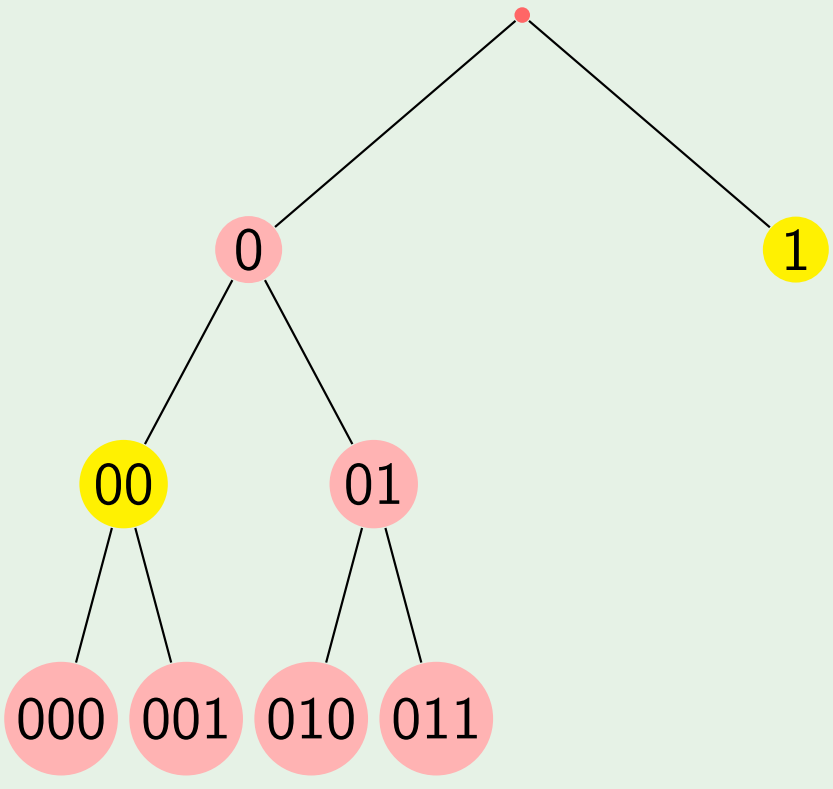
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

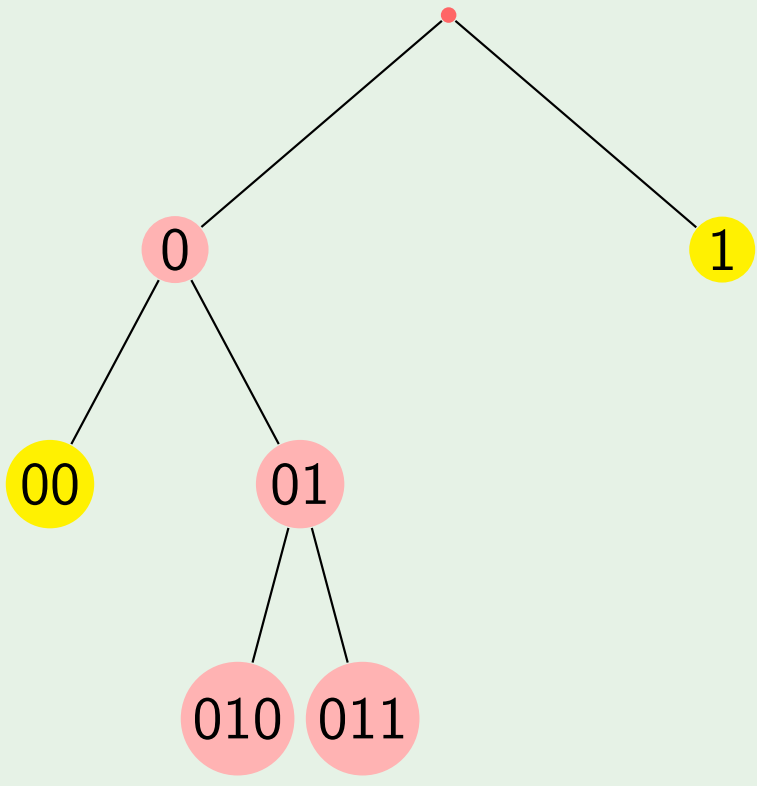
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

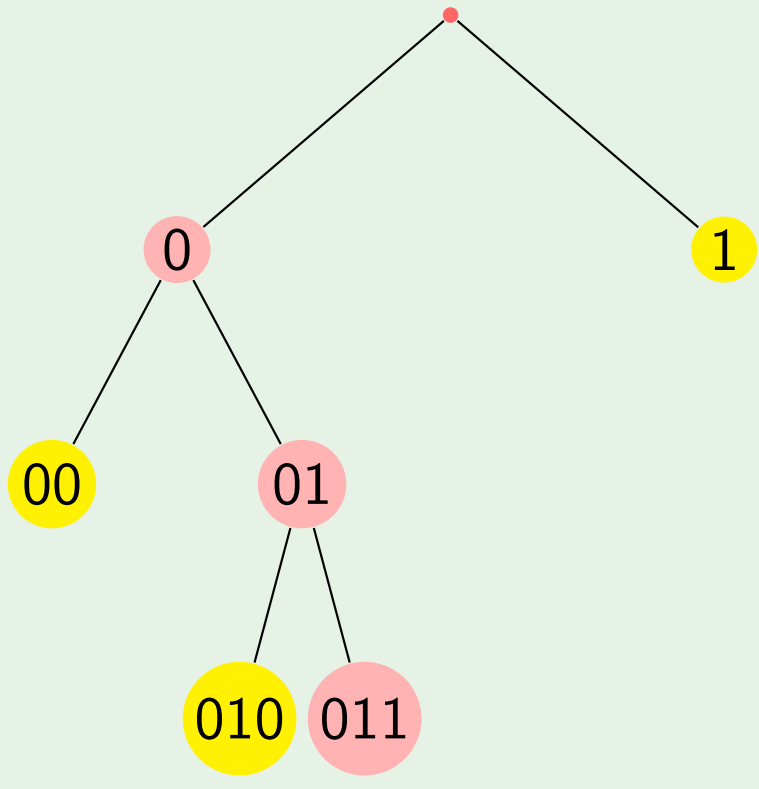
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

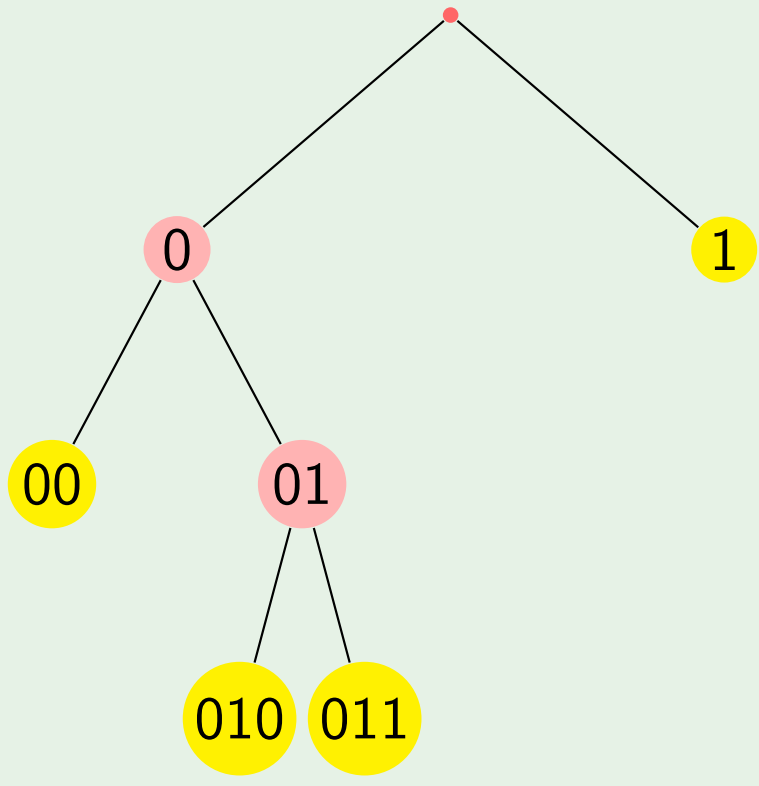
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Kraftova nerovnost pro jednoznačně dekódovatelné kódy

Věta (McMillan, 1956)

Délky slov $\ell_1, \ldots, \ell_m$ libovolného **jednoznačně dekódovatelného** d-znakového kódu splňují nerovnost

$$\sum_{i=1}^{m} d^{-\ell_i} \leqslant 1.$$

Obráceně, splňují-li $\ell_1, \ldots, \ell_m \in \mathbb{N}$ tuto nerovnost, potom existuje **jednoznačně dekódovatelný** d-znakový kód s délkami slov $\ell_1, \ldots, \ell_m$.

Hledáme optimální kód

- Kraftova nerovnost umožňuje konstrukci instantního kódu pomocí zadaných délek kódových slov $\ell_1, \ldots, \ell_m$
- hledání kódu minimální střední délky tak lze formulovat jako optimalizační úlohu
- pro pravděpodobnostní funkci p_X na m-prvkové množině Λ a $d \in \mathbb{N}$ tak hledáme minimum funkce

$$\bar{L}(\ell_1, \ldots, \ell_m) = \sum_{i=1}^{m} p_X(x_i)\ell_i$$

na množině

$$\left\{ (\ell_1, \ldots, \ell_m) \in \mathbb{N}^m \;\middle|\; \sum_{i=1}^{m} d^{-\ell_i} \leqslant 1 \right\}$$

Hledáme optimální kód (pokr.)

- zanedbejme celočíselné omezení kladené na $(\ell_1, \ldots, \ell_m)$ a předpokládejme v Kraftově nerovnosti rovnost:

$$\sum_{i=1}^{m} d^{-\ell_i} = 1$$

- potom snadno nalezneme vázaný extrém funkce $\bar{L}$ metodou Lagrangeových multiplikátorů:

$$\ell'_i = -\log_d p(x_i), \quad i = 1, \ldots, m$$

- v bodě $(\ell'_1, \ldots, \ell'_m)$ je hodnota minima rovna Shannonově entropii:

$$\bar{L}(\ell'_1, \ldots, \ell'_m) = \sum_{i=1}^{m} p_X(x_i)\ell'_i = H_d(X)$$

Hledáme optimální kód (pokr.)

- nalezli jsme pouze vektor $(\ell'_1, \ldots, \ell'_m) \in \mathbb{R}^m$!
- pro každou d-adickou p_X platí $(\ell'_1, \ldots, \ell'_m) \in \mathbb{N}^m$
- můžeme se pokusit najít instantní kód, jehož délky slov by se "příliš nelišily" od $(\ell'_1, \ldots, \ell'_m)$, tato úloha je však obtížně řešitelná a proto se prosadily jiné metody

Klasické kódovací algoritmy

1. Shannonovo kódování
2. Fanovo kódování
3. Huffmanovo kódování
4. LZ algoritmy

Žádný způsob kódování nepřekoná mez danou entropií!

Dolní mez komprese

Věta

- Střední délka $L(C)$ libovolného jednoznačně dekódovatelného d-znakového kódu C pro náhodnou veličinu X splňuje nerovnost
$$L(C) \geqslant H_d(X).$$
- Rovnost
$$L(C) = H_d(X)$$
platí právě tehdy, když $p_X(x) = d^{-\ell(C(x))}$.

Nutné podmínky optimality

Optimální kód zatím neumíme nalézt. Ovšem snadno odvodíme, jaké musí splňovat podmínky.

Věta

Nechť C je optimální binární instantní kód pro veličinu X. Platí:

1. Pokud $p_X(x) > p_X(y)$, pak $\ell(C(x)) \leqslant \ell(C(y))$.
2. V kód. stromě pro C má každý list přiřazeno kódové slovo.
3. Dvě nejméně pravděpodobná kódová slova mají stejnou délku.

Shannonovo kódování

Vstup: zdroj (Λ, p), kde $\Lambda = \{x_1, \ldots, x_n\}$, $p = (p_1, \ldots, p_n)$
Výstup: kód $C_S : \Lambda \to \{0, 1, \ldots, d-1\}^*$

- bez újmy na obecnosti předpokládáme $p_i > 0$
- položme

$$\ell_i = \left\lceil \log_d p_i^{-1} \right\rceil,$$

tato čísla splňují Kraftovu nerovnost:

$$\sum_{i=1}^{n} d^{-\left\lceil \log_d p_i^{-1} \right\rceil} \leqslant \sum_{i=1}^{n} d^{-\log_d p_i^{-1}} = \sum_{i=1}^{n} p_i = 1$$

- slova kódu C_S nalezneme pomocí konstrukce instantního kódu se zadanými délkami ℓ_i

Meze Shannonova kódování

Věta

Platí $H_d(X) \leqslant L(C_S) < H_d(X) + 1$.

Důkaz.

Snadno plyne z nerovnosti

$$\log_d p_i^{-1} \leqslant \left\lceil \log_d p_i^{-1} \right\rceil < \log_d p_i^{-1} + 1.$$

Poznámka. Střední délka optimálního kódu C^* tedy splňuje

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

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování
- Zdroje s pamětí

Úvod

- Huffman nalezl v r.1951 optimální kód C_H, tedy instantní kód minimalizující střední délku na množině všech jednoznačně dekódovatelných kódů
- výsledný kód není určen jednoznačně (např. bitovou inverzí získáme jiný optimální kód)
- jednoznačně není určena ani délka kódových slov
- aplikace
  - závěrečné zpracování formátů JPEG, MP3, DEFLATE, PKZIP
  - předzpracování souboru pro aritmetické kódování

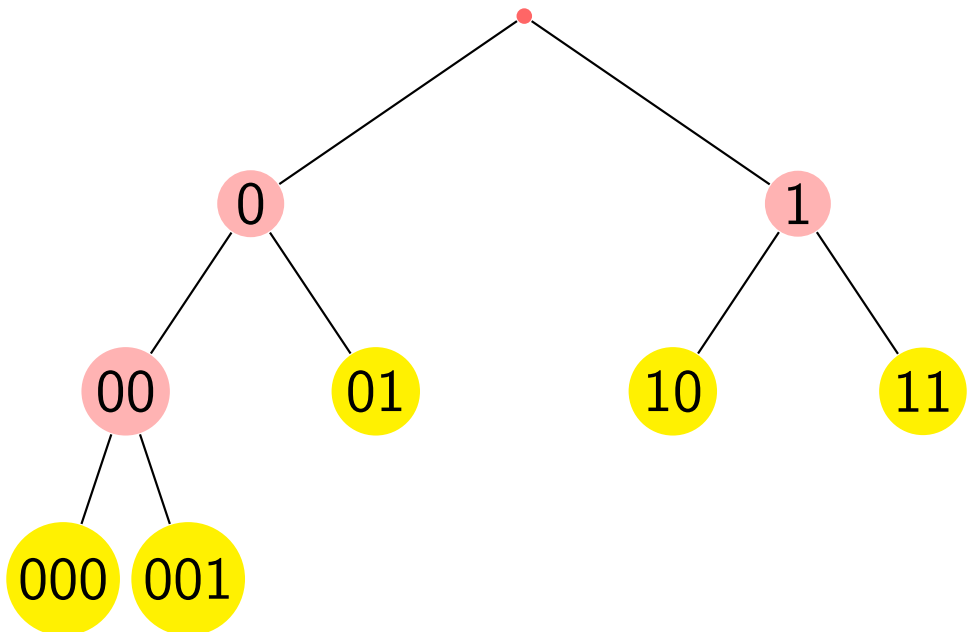
Algoritmus

Vstup: inf. zdroj (Λ, p), kde $\Lambda = \{x_1, \ldots, x_n\}$, $p = (p_1, \ldots, p_n)$
Výstup: kód $C_H : \Lambda \to \{0,1\}^*$

1. Z prvků množiny Λ vytvoř množiny $S_1 = \{x_1\}, \ldots, S_n = \{x_n\}$, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a uvažuj informační zdroj $(\mathcal{S}, p)$
2. Dokud $\mathcal{S}$ není jednoprvková:
   - 2-1 najdi množiny S_i, S_j, $i \neq j$ s nejnižšími pravděpodobnostmi p_i, p_j
   - 2-2 prvkům z S_i připiš bit 0, prvkům z S_j připiš bit 1 (bity připisujeme na **začátek** kódového slova)
   - 2-3 polož $\mathcal{S} := \mathcal{S} \setminus \{S_i, S_j\}$
   - 2-4 polož $\mathcal{S} := \mathcal{S} \cup \{S_i \cup S_j\}$ a $p(S_i \cup S_j) := p(S_i) + p(S_j)$
3. Každému $x_i \in \Lambda$ přiřaď slovo $C_H(x_i)$, které vzniklo postupným připisováním bitů

Příklad

x	$p(x)$	x	$p(x)$	x	$p(x)$	x	$p(x)$	kód
1	0.25	$(4,5)$	0.30	$(2,3)$	0.45	$(1,4,5)$	0.55	01
2	0.25	1	0.25	(4,5)	0.30	(2,3)	0.45	10
3	0.20	2	0.25	1	0.25			11
4	0.15	3	0.20					000
5	0.15							001



Huffmanův vs. Shannonův kód

Srovnání délek kód. slov:

Příklad

Informační zdroj: $\Lambda = \{x_1, x_2\}$, $p(x_1) = 0.9999, p(x_2) = 0.0001$.

- Shannonův kód obsahuje slova délky 1 a 14
- Huffmanův kód obsahuje slova délky 1 a 1

Příklad

Informační zdroj: $\Lambda = \{x_1, x_2, x_3, x_4\}$,

$$p(x_1) = p(x_2) = 3^{-1}, p(x_3) = 4^{-1}, p(x_4) = 12^{-1}.$$

- Huffmanův kód má slova délek $(2, 2, 2, 2)$ nebo $(1, 2, 3, 3)$
- Shannonův kód dává pro x_3 slovo délky 2

Optimalita Huffmanova kódování

Věta

Nechť C_H je Huffmanův kód a C je libovolný jednoznačně dekódovatelný kód pro náhodnou veličinu X. Potom platí:

- C_H je optimální, neboli $L(C_H) \leqslant L(C)$
- $H(X) \leqslant L(C_H) < H(X) + 1$

Vlastnosti Huffmanova kódování

Výhody

- minimální střední délka
- snadná implementace
- není patentová ochrana

Nevýhody

- vstupem je informační zdroj, což vyžaduje načtení celého souboru dat kvůli výpočtu četností jednotlivých symbolů (řešení: **adaptivní Huffmanovo kódování**)
- ke kódu je nutno připojit **kódovací tabulku**

Připojení kódovací tabulky

Příklad

Zpráva obsahuje 10^4 znaků ($=10^4$ bajtů) z abecedy $\Lambda = \{a, \ldots, e\}$ s pravděpodobnostmi

$$p_X(a) = 0.35,\ p_X(b) = p_X(c) = 0.17,\ p_X(d) = 0.16,\ p_X(e) = 0.15.$$

Zkomprimovaná zpráva má

$$10^4 L(C_H) = 10^4 \cdot 2.3 = 23\,000 \text{ bitů} = 2\,875 \text{ bajtů}$$

Kompresní poměr 28.75 % je tak připojením kódovací tabulky navýšen jen minimálně.

Ovšem $H(X) = 2.23284 < 2.3 = L(C_H)$, rozdíl je asi 3 %. Entropii se můžeme dále přiblížit **blokovým kódováním**.

Blokové kódování

Věta (Shannonova věta o zdrojovém kódování)

Nechť $(X_n)_{n\in\mathbb{N}}$ je **bezpaměťový** zdroj s rychlostí entropie $H((X_n)_{n\in\mathbb{N}}) = H(X_1)$. Potom Huffmanův nebo Shannonův kód C^n pro n-tice znaků z abecedy Λ splňuje

$$\lim_{n\to\infty} \frac{L(C^n)}{n} = H(X_1).$$

Příklad

$\Lambda = \{a, b, c\}$, $p_X(a) = 0.1$, $p_X(b) = 0.2$, $p_X(c) = 0.7$
$L(C_H) = 0.1 \cdot 2 + 0.2 \cdot 2 + 0.7 \cdot 1 = 1.3 > H(X) = 1.15678$

Lze dosáhnout maximálně 11% úspory **blokovým kódováním**.

Blokové kódování-příklad

Příklad(pokr.)

Uvažujme Huffmanův blokový kód pro **dvojice** znaků z Λ. Zdrojová abeceda je nyní Λ^2 a pravděpodobnosti stanovíme díky nezávislosti jako

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y), \quad (x, y) \in \Lambda^2$$

Potom pro kód C_H^2 platí

$$\frac{L(C_H^2)}{2} = \frac{2.33}{2} = 1.165$$

V porovnání s kódem C_H, pro který platí $L(C_H) = 1.3$, tak byla zvýšena komprese o 10 %, ovšem za cenu připojení větší kódovací tabulky a prodloužení odezvy při kódování.

Shrnutí

- dolní mez komprese je rovna **entropii** informačního zdroje
- **Huffmanovo kódování** je optimální kompresní algoritmus pro jednu diskrétní náhodnou veličinu
- **blokové kódování** umožňuje asymptoticky optimálně zkomprimovat libovolný **bezpaměťový** zdroj informace
- ovšem kódování **není jednoprůchodové**, je nutno nejdříve načíst celý vstupní soubor
- umíme optimálně komprimovat zdroje s pamětí?

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování
- Zdroje s pamětí

Zdroj s pamětí

- bezpaměťový zdroj je posloupnost $(X_n)_{n\in\mathbb{N}}$ nezávislých a stejně rozdělených náh. veličin nad konečnou abecedou Λ
- většina informačních zdrojů (texty, bitmapy) však vykazuje velmi silnou závislost mezi sousedními znaky X_i a X_{i+1} či sousedními řetězci $X_{i-k} \ldots X_i$ a X_{i+1} (**markovské zdroje**)
- pro zdroje s pamětí nemusí být Huffmanovo kódování optimální
- komprimovatelnost stacionárního zdroje s pamětí určuje **rychlost entropie**

Příklad zdroje s pamětí

Entropie českého textu - viz tato studie

Předpokládejme, že náhodně vybraný vzorek dlouhého českého textu (např. beletrie) tvoří **stacionární zdroj informace** nad abecedou

$$\Lambda = \{\mathrm{a}, \ldots, \mathrm{z}, \acute{\mathrm{a}}, \ldots, \check{\mathrm{z}}, \mathrm{ch}\}.$$

Odhadneme jeho rychlost entropie $H((X_n)_{n\in\mathbb{N}})$. Nejprve stanovíme entropii rovnoměrného modelu:

$$\log|\Lambda| = 5.39 \text{ bitu}.$$

Vezmeme-li v úvahu četnosti výskytu jednotlivých písmen, dostaneme entropii bezpaměťového zdroje:

$$H(X) = 4.72 \text{ bitu}.$$

Příklad zdroje s pamětí

Entropie českého textu - pokr.

Pokud uvažujeme bigramy (markovský zdroj), dostaneme entropii

$$H(X_2|X_1) = 3.69 \text{ bitu.}$$

Při použití trigramů (2-markovský zdroj)

$$H(X_3|X_2, X_1) = 3.18 \text{ bitu.}$$

Další přiblížení jsou výpočetně náročná. Rychlost entropie lze odhadnout jako

$$H((X_n)_{n\in\mathbb{N}}) \approx 2.07 \text{ bitu.}$$

Závěr: písmeno v průměrném textu nese asi 2 bity informace. Pro abecedu s mezerou je výsledek téměř stejný.

Jak komprimovat zdroje s pamětí?

- autoři Lempel a Ziv publikovali 2 základní varianty algoritmu LZ77 a LZ78 (1977-1978)
- algoritmus má mnoho různých variací, jako je např. Lempel-Ziv-Welsch LZW (`compress` v Unixu, ZIP, RAR, komprese v modemech, GIF, PDF)
- jde o třídu adaptivních kompresních algoritmů se **slovníkem**
- nevyžaduje znalost rozdělení zdroje, kódování probíhá **jednoprůchodově**

LZ78 (stromová verze LZ)

- řetězec $x_1 \dots x_n \in \Lambda^n$ je sekvenčně testován na výskyt nejkratších řetězců, které se nevyskytly v předchozím kroku
- každý takový řetězec je označen a uložen do slovníku
- díky minimalitě ukládaného řetězce jsou jeho prefixy již ve slovníku: řetězec $x_i \dots x_k$ byl uložen do slovníku před $x_i \dots x_k x_{k+1}$

Kód tvoří posloupnost dvojic (U_k, x_k), kde

- x_k je poslední znak řetězce $x_i \dots x_\ell x_k$,
- U_k je ukazatel na odpovídající prefix $x_i \dots x_\ell$.

Ilustrace LZ78

Příklad

$\Lambda = \{A, B\}$, řetězec $ABBABBABBBAABABAA$
Dostaneme následující řetězce:

$$A, B, BA, BB, AB, BBA, ABA, BAA$$

Výsledný kód:

$$0A\ 0B\ 2A\ 2B\ 1B\ 4A\ 5A\ 3A$$

Efektivita LZ78

Nechť $c(n)$ je počet slov ve slovníku pro zprávu $x_1 \dots x_n \in \{0,1\}^n$.

- kód se skládá z $c(n)$ dvojic (ukazatel,poslední bit)
- potřebujeme tedy celkem nejvýše

$$c(n)(\log c(n) + 1) \text{ bitů}$$

- pro stacionární ergodický zdroj informace $(X_n)_{n\in\mathbb{N}}$ je LZ78 **asymptoticky optimální komprimátor**, neboť platí

$$\frac{c(n)(\log c(n) + 1)}{n} \to H((X_n)_{n\in\mathbb{N}})$$

s pravděpodobností 1

Část V

Informační kanály

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování
- Zdroje s pamětí

Motivace

Příklad (Vajda, 2004)

Holub má přepravit 1 bit informace (výhra/prohra bitvy). Jaká je kapacita použitého informačního kanálu v případě, že

1. bude holub sežrán sokolem s pravděpodobností 0.2,
2. holub se vyhne sokolům, ale s pravděpodobností 0.2 je odchycen špiónem, který mu zprávu změní na opačnou.

Informační kanál

Definice

Informační kanál je trojice $\mathcal{K} = (\Lambda, \mathbf{P}, \Omega)$, kde

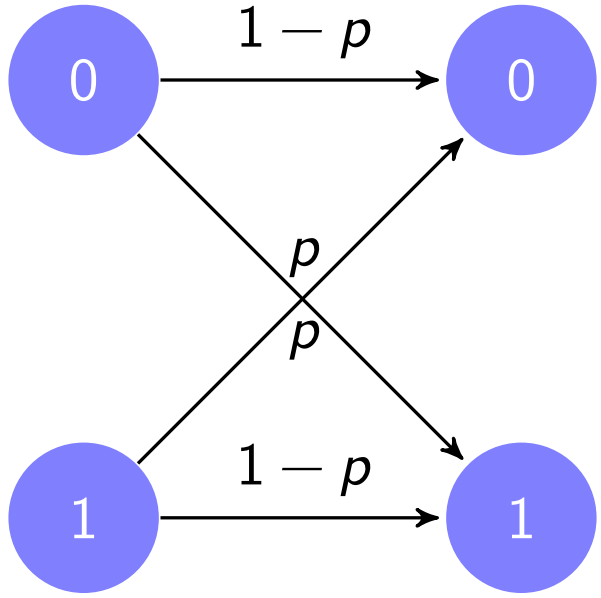
- Λ je m-prvková vstupní abeceda
- Ω je n-prvková výstupní abeceda
- $\mathbf{P}$ je matice $m \times n$ podmíněných pravděpodobností:

$$\mathbf{P} = \begin{pmatrix} p_{Y|X}(y_1|x_1) & p_{Y|X}(y_2|x_1) & \dots & p_{Y|X}(y_n|x_1) \\ p_{Y|X}(y_1|x_2) & p_{Y|X}(y_2|x_2) & \dots & p_{Y|X}(y_n|x_2) \\ & \dots\dots\dots\dots\dots\dots & & \\ p_{Y|X}(y_1|x_m) & p_{Y|X}(y_2|x_m) & \dots & p_{Y|X}(y_n|x_m) \end{pmatrix}$$

Typicky $\Lambda = \Omega = \{0, 1\}$.

Příklad: binární symetrický kanál

$\Lambda = \Omega = \{0, 1\}$



$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

- Model komunikace
- Kapacita kanálu
- Shannonova věta o kapacitě

Přenositelnost informace kanálem

Vstup: informační zdroj X s pravděpodobnostmi

$$\big(p_X(x_1), \ldots, p_X(x_m)\big)$$

Výstup: informační zdroj Y s pravděpodobnostmi

$$\big(p_Y(y_1), \ldots, p_Y(y_n)\big) = \big(p_X(x_1), \ldots, p_X(x_m)\big) \cdot \mathbf{P}$$

- pravděpodobnosti $p_{X|Y}(x_i|y_j)$ lze určit pomocí Bayesova vzorce
- přenositelnost zdroje X popisuje **vzájemná informace**

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

Definice a vlastnosti

Informační kapacita

Volba pravděpodobností p_X je na uživateli, neboť ty odpovídají četnostem použitých znaků kanálové abecedy Λ.

Definice

Informační kapacita kanálu $\mathcal{K} = (\Lambda, \mathbf{P}, \Omega)$ je

$$C(\mathcal{K}) = \sup_{p_X} I(X; Y).$$

- $I(X; Y)$ je spojitou funkcí p_X na kompaktní množině Δ_m, supréma se nabývá pro nějaké $p_X \in \Delta_m$
- vztah pro $I(X; Y)$ je symetrický:

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

Vlastnosti kapacity

Věta

Nechť $\mathcal{K} = (\Lambda, \mathbf{P}, \Omega)$ je informační kanál. Potom:

- $C(\mathcal{K}) \geqslant 0$
- $C(\mathcal{K}) \leqslant \log |\Lambda|$
- $C(\mathcal{K}) \leqslant \log |\Omega|$

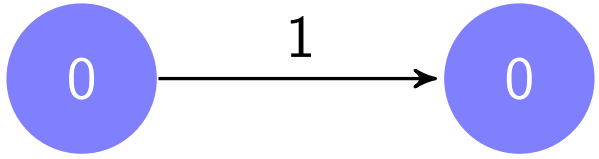
Výpočet kapacity

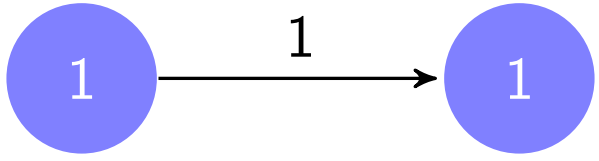
- hledáme maximum rozdílu $H(Y) - H(Y|X)$ pro X
- výpočet je snadný pro kanály, v nichž se pravděpodobnostní funkce $p_{Y|X}(\cdot|x_i)$, $p_{Y|X}(\cdot|x_j)$ liší pouze permutací odpovídajících pravděpodobností
- odpovídající matice přechodu $\mathbf{P}$ je tak tvořena permutacemi jednoho řádku
- to znamená, že $H(Y|x_i) = H(Y|x_j)$ pro každé $x_i, x_j \in \Lambda$ a proto

$$H(Y|X) = \sum_{x_k \in \Lambda} p_X(x_k) H(Y|x_k) = H(Y|x_i)$$

Binární bezšumový kanál

$\Lambda = \Omega = \{0, 1\}$

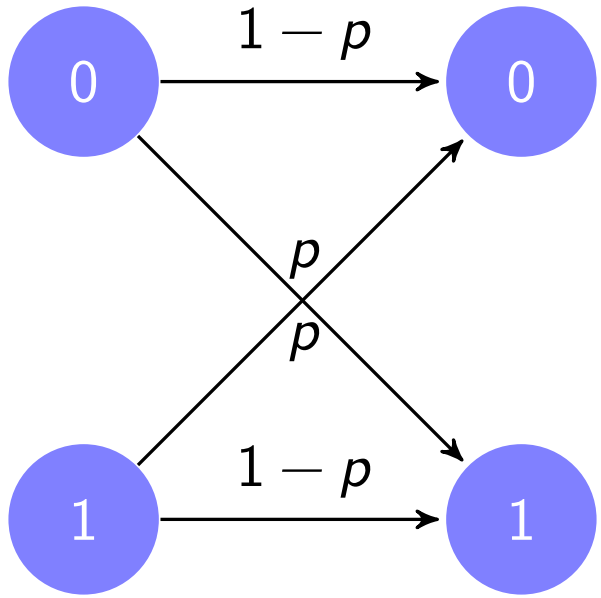




Kapacita $C(\mathcal{K}) = 1$

Binární symetrický kanál

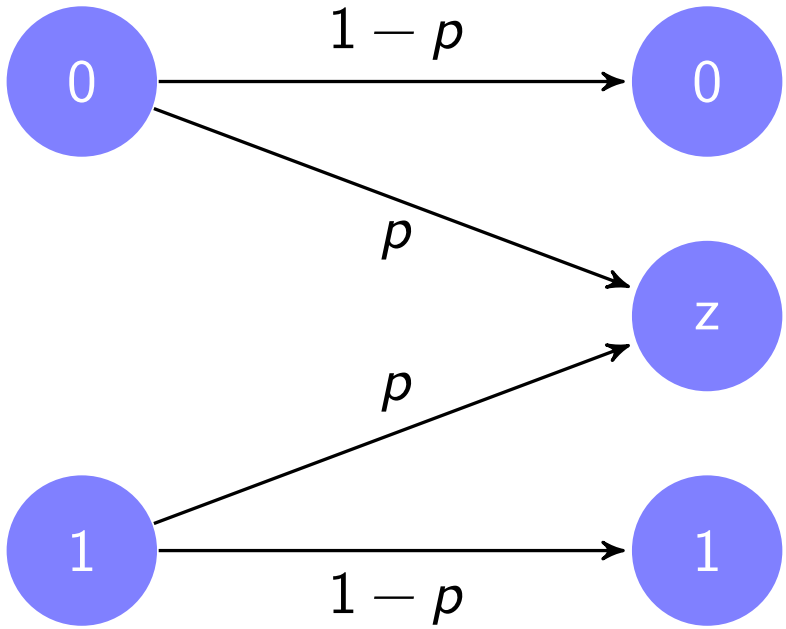
$\Lambda = \Omega = \{0, 1\}$



Kapacita $C(\mathcal{K}) = 1 - H(p, 1-p)$

Binární kanál se zámlkou

$\Lambda = \{0, 1\}, \Omega = \{0, 1, z\}$



Kapacita $C(\mathcal{K}) = 1 - p$

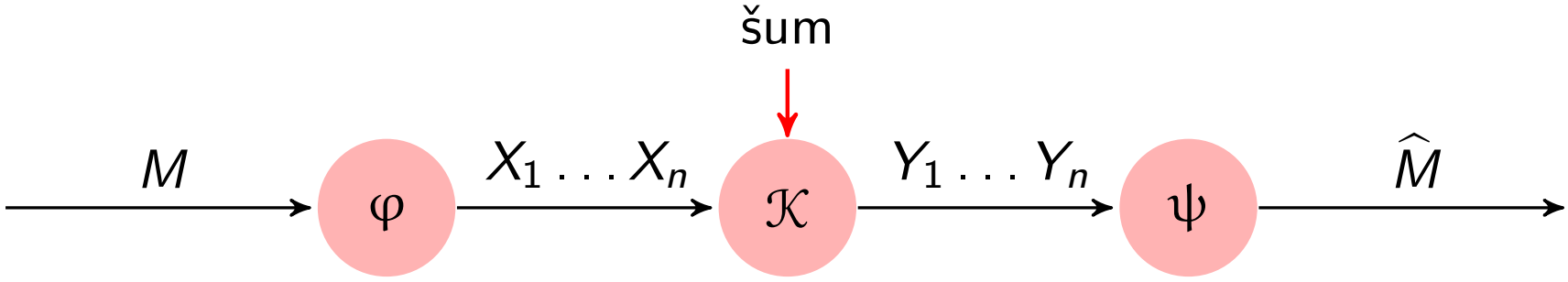
Zámlka lepší než záměna

Věta

Kapacita binárního kanálu s pravděpodobností zámlky $2p$ $(0 < p < \frac{1}{2})$ je větší než kapacita binárního symetrického kanálu s pravděpodobností záměny p.

- Model komunikace
- Kapacita kanálu
- Shannonova věta o kapacitě

Schéma komunikace



- **náhodnost** spočívá v přítomnosti šumu a zdrojové zprávy M
- **uživatel** volí kodér φ a dekodér ψ, délku kód. slova n

Složky modelu komunikace

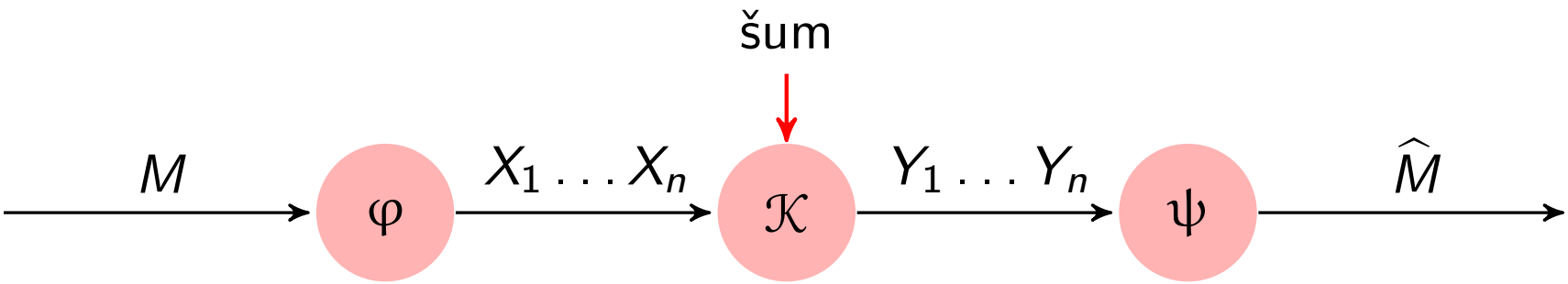
1. **zdroj** je náhodná veličina M s rovnoměrným rozdělením na konečné množině zpráv $\mathcal{M}$
2. **kodér** je zobrazení
$$\varphi : \mathcal{M} \to \Lambda^n,$$
které každé zdrojové zprávě $i \in \mathcal{M}$ přiřazuje vstupní kódové slovo $\varphi(i) = x_1 \dots x_n \in \Lambda^n$
3. **bezpaměťový informační kanál** $\mathcal{K} = (\Lambda, \mathbf{P}, \Omega)$, ten splňuje
$$p_{Y_k|X_1 \dots X_k Y_1 \dots Y_{k-1}} = p_{Y_k|X_k}$$
4. **dekodér** je zobrazení
$$\psi : \Omega^n \to \mathcal{M},$$
které každému výstupu kanálu $y_1 \dots y_n \in \Omega^n$ přiřazuje odhad $\psi(y_1 \dots y_n) \in \mathcal{M}$ zaslané zprávy M

Kanálové kódování

Definice

$(|\mathcal{M}|, n)$-kódovací schéma pro kanál $(\Lambda, \mathbf{P}, \Omega)$ je tvořeno

- množinou $\mathcal{M}$ stejně pravděpodobných zpráv,
- kodérem φ, který každé zprávě $i \in \mathcal{M}$ přiřazuje slovo $x_1 \dots x_n \in \Lambda^n$,
- dekodérem ψ, který každému výstupu $y_1 \dots y_n \in \Omega^n$ přiřazuje odhad původní zaslané zprávy z $\mathcal{M}$.



Rychlost přenosu

Definice

Rychlost $(|\mathcal{M}|, n)$-kódovacího schématu je

$$R = \frac{\log |\mathcal{M}|}{n} \quad \text{bit/znak kanálové abecedy.}$$

- R udává rychlost komunikace daným kanálem
- jiné jednotky: bit/1 přenos kanálem
- chceme $R \to$ MAX při současné minimální chybě!

Pravděpodobnosti chyb

Definice

Pro libovolné $(|\mathcal{M}|, n)$-kódovací schéma pro kanál $\mathcal{K}$ definujeme:

- podmíněná pravděpodobnost chyby pro $i \in \mathcal{M}$ je

$$\lambda_i = P[\psi(Y_1 \dots Y_n) \neq i \,|\, (X_1 \dots X_n) = \varphi(i)]$$

- maximální pravděpodobnost chyby je

$$\lambda^{(n)} = \max_{i \in \mathcal{M}} \lambda_i$$

- průměrná pravděpodobnost chyby je $p_e^{(n)} = \frac{1}{|\mathcal{M}|} \sum_{i=1}^{|\mathcal{M}|} \lambda_i$

Hlavní cíl kanálového kódování

Hledáme $(|\mathcal{M}|, n)$-kódovací schéma minimalizující maximální pravděpodobnost chyby $\lambda^{(n)}$ pro zadanou rychlost R.

Jak navrhnout dekodér $\psi : \Omega^n \to \mathcal{M}$, aby tohoto cíle dosáhl? Položme

$$\psi(y_1 \dots y_n) = \arg\max_{i \in \mathcal{M}} p_{Y_1 \dots Y_n | X_1 \dots X_n}(y_1 \dots y_n | \varphi(i))$$

pro každé $y_1 \dots y_n \in \Omega^n$.

Co říká Shannonova věta?

Věta

Nechť $\mathcal{K}$ je informační kanál. Potom:

- Lze nalézt kód o rychlosti $R < C(\mathcal{K})$ takový, že chyba $\lambda^{(n)}$ bude libovolně malá pro $n \to \infty$.
- Naopak, kód o rychlosti přenosu $R > C(\mathcal{K})$ má nutně chybu $\lambda^{(n)} > 0$ pro $n \to \infty$.

Kapacita kanálu $C(\mathcal{K})$ je tak mezí spolehlivé komunikace **nezávisle** na použitém kódovacím algoritmu!

Komentář k Shannonově větě

- důkaz je konstruktivní
- ovšem je nutno prohledat všechny kodéry, těch je $O(2^{2^n})$
- pro velká n má navíc systém pomalou odezvu a složitá pravidla
- pro praktické použití se tedy nehodí...
- honba za kódy, jejichž rychlost se přibližuje kapacitě kanálu, začala v 50.letech minulého století a pokračuje dodnes

Část VI

Dodatek

Značení a konvence

- log znamená $\log_2$
- pokud A je konečná množina, pak $|A|$ značí počet prvků v A
- pro libovolné $x \in \mathbb{R}$ značí symbol $\lceil x \rceil$ nejmenší celé číslo y takové, že $x \leqslant y$

Jensenova nerovnost

Věta

Nechť $f\colon (a,b) \to \mathbb{R}$ je konkávní funkce. Je-li $t_1, \ldots, t_n \in (a,b)$, potom pro všechna $\alpha_1, \ldots, \alpha_n \in \langle 0,1 \rangle$ splňující $\sum_{i=1}^{n} \alpha_i = 1$ platí

$$f\left(\sum_{i=1}^{n} \alpha_i t_i\right) \geqslant \sum_{i=1}^{n} \alpha_i f(t_i).$$

Pokud je f ryze konkávní, potom rovnost v Jensenově nerovnosti implikuje $t_1 = \cdots = t_n$.

Pravděpodobnostní interpretace

Je-li X náhodná veličina s konečným výběrovým prostorem $\mathcal{X} \subseteq \mathbb{R}$ a f je konkávní funkce na nějakém intervalu obsahujícím $\mathcal{X}$, potom

$$f(EX) \geqslant E(f(X)).$$

Kam dále v teorii informace?

- ergodická teorie stacionárních náhodných procesů nad konečnou abecedou jako základní rámec pro moderní kompresní algoritmy (LZW)
- kanálové kódování jako algebraická teorie pro spolehlivou komunikaci
- ztrátová komprese
- Gaussovské informační kanály
- kvantová teorie informace

Literatura


Thomas M. Cover and Joy A. Thomas.
Elements of Information Theory.
Wiley-Interscience [John Wiley & Sons], NJ, 2006.


S.M. Moser.
Information Theory
http://moser-isi.ethz.ch/scripts.html


I. Vajda.
Teorie informace.
Vydavatelství ČVUT, 2004.


A. Drozdek.
Elements of Data Compression.
Brooks/Cole, 2002.