Úvod do teorie kódování

Matematické základy komprese a digitální komunikace

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

upravil Mirko Navara
http://cmp.felk.cvut.cz/~navara/

12. 1. 2017

Part 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, do 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.
A mathematical theory of communication.
Bell System Tech. J., 27:379–423, 623–656, 1948.

Model komunikace podle Shannona

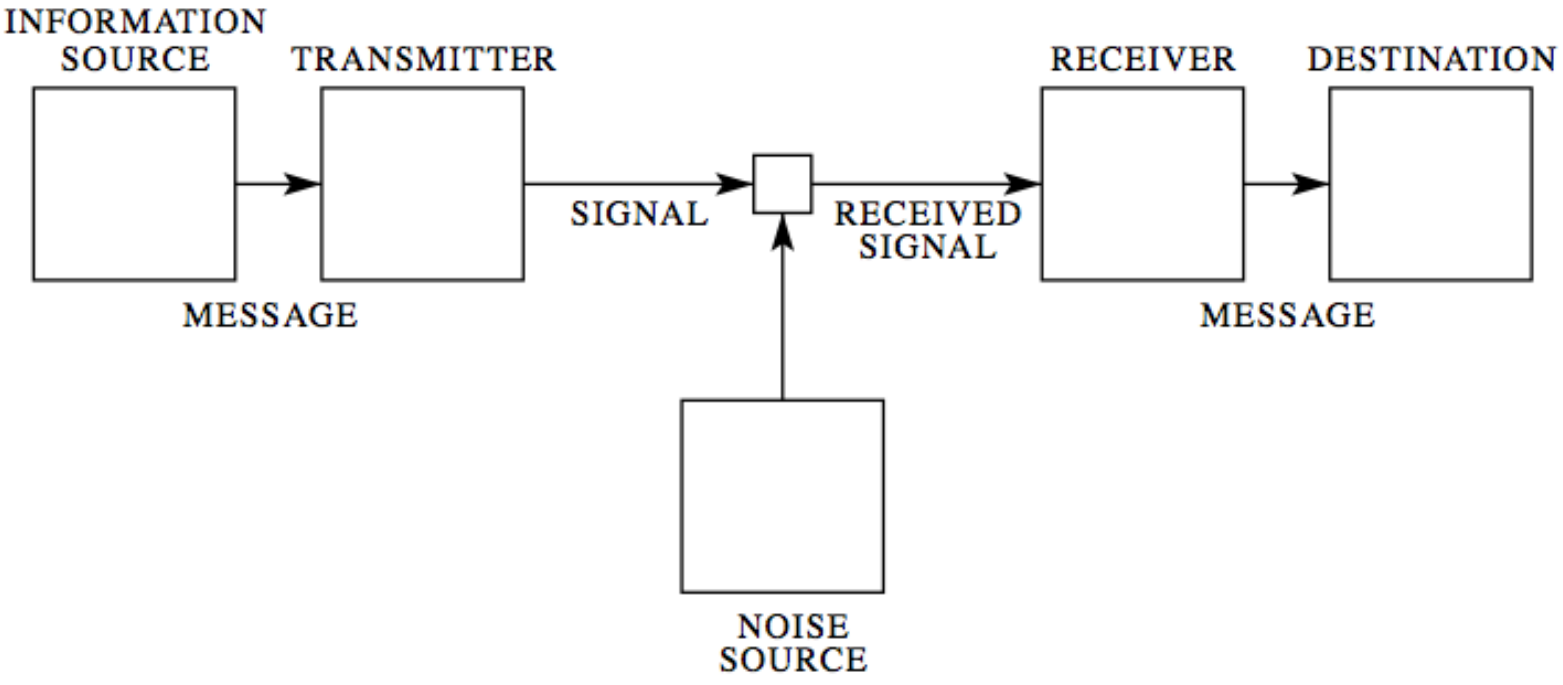


Fig. 1 — Schematic diagram of a general communication system.

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

kódování $\rightarrow$ přenos $\rightarrow$ dekódování
Nelze je posuzovat odděleně!

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

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

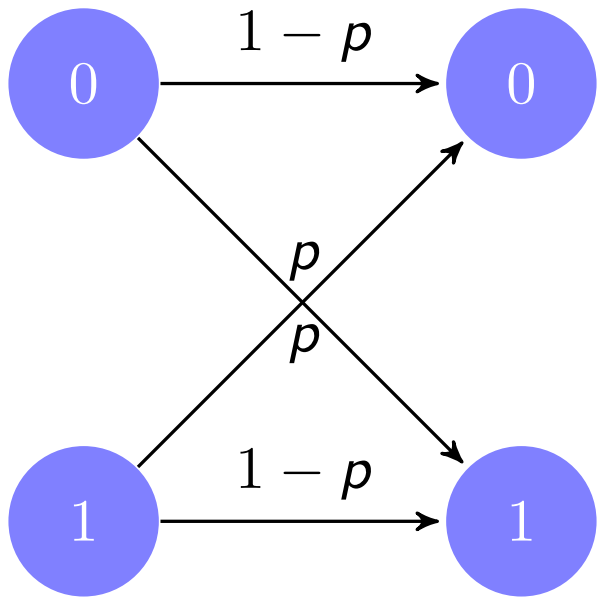
- střední délka kódu s 3-bitovými délkami kódových slov je 3
- uvažujme následující kód s délkami slov $\ell_i = -\log p_i$

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 $y_1 \cdots y_{2n+1}$

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!

Part II

Charakteristiky informace

- Entropie
- Vzájemná informace

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 $k \geq 1$, potom $I(2^k) = k$ udává počet bitů, kterými lze zakódovat znak z Λ
- pokud neexistuje $k \geq 1$ 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

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

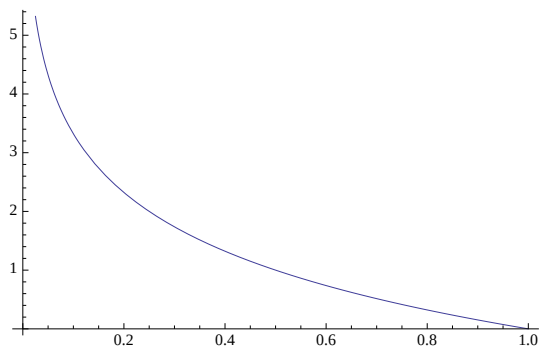
Entropie

Definice (Shannon)

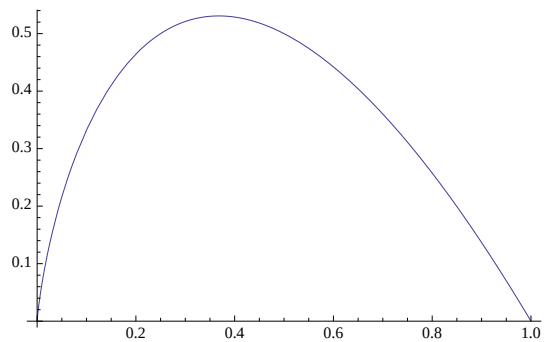
Nechť X je náhodná veličina s hodnotami v Λ a pravděpodobnostní funkcí p_X. Entropie náhodné veličiny X je

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

kde užíváme konvenci $0 \log 0 = 0$.



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



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

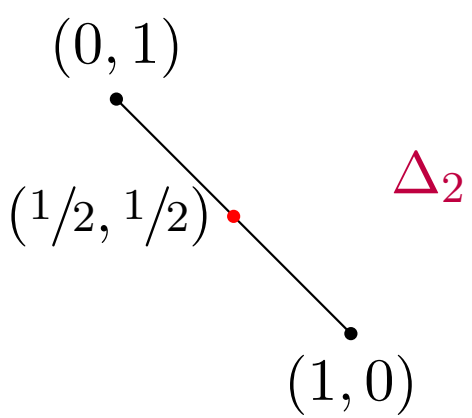
Entropie jako funkce

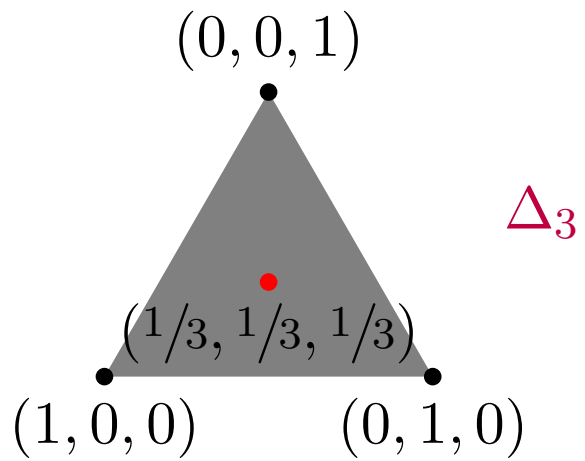
Nechť $|\Lambda| = n$. Entropii lze chápat jako reálnou funkci

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

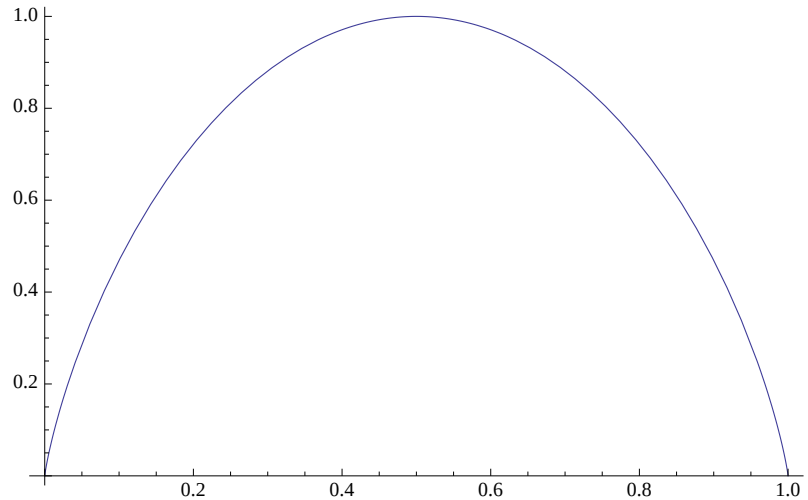
kde Δ_n je pravděpodobnostní n-simplex, což je množina všech pravděpodobnostních funkcí na Λ:

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





Entropie jako funkce (pokr.)



Entropie na Δ_2

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 $\{\{1, 2\}, 3, 4, 5\}$ 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

$$H_d(X) = \log_d 2 \cdot H(X) = \frac{H(X)}{\log_2 d}.$$

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 $H(X) \geq 0$. 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 $H(X) \leq \log |\Lambda|$. Rovnost $H(X) = \log |\Lambda|$ nastane právě tehdy, když má X rovnoměrné rozdělení.

Důkaz.

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

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

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

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. $0 \leq I(X;Y) \leq H(X)$
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 $t > 0$:

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

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

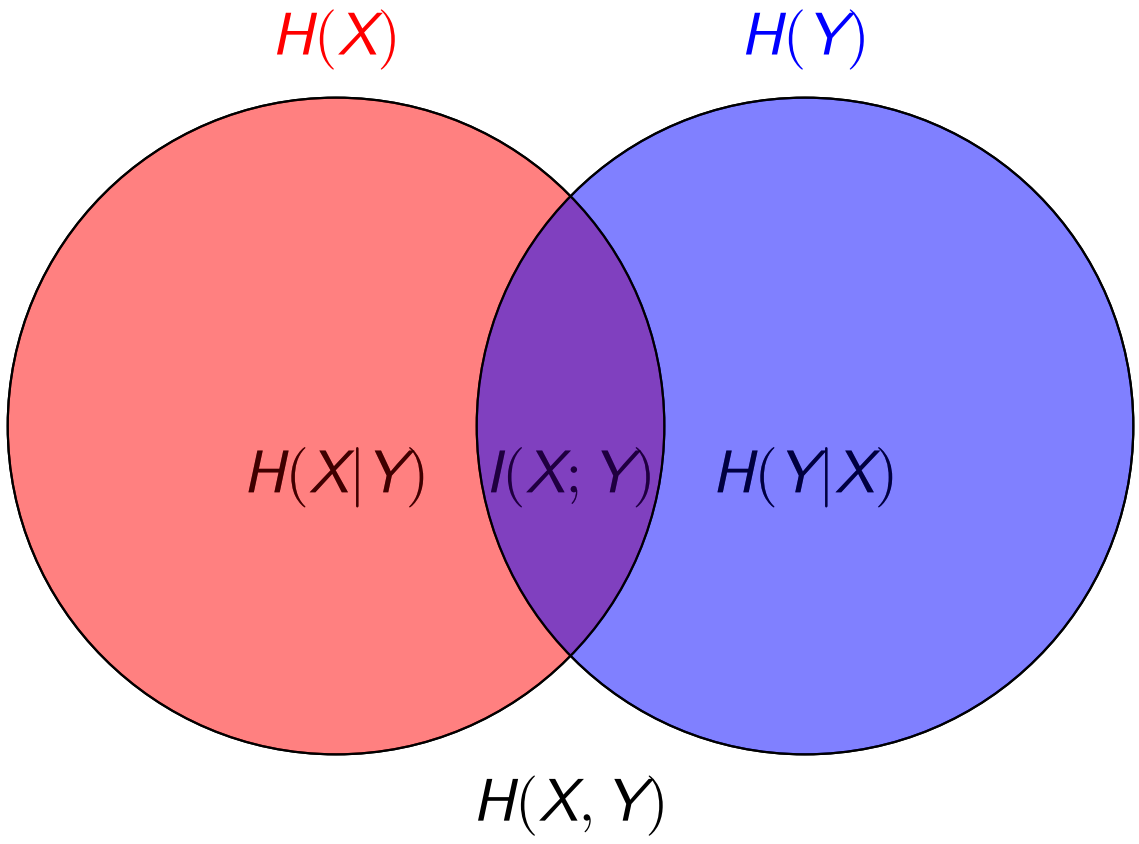
$$p_{X|Y}(1|0) = p_{X|Y}(0|1) = 1.$$

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.

Part III

Bezšumový přenosový kanál bez paměti

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování

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 $n \in \mathbb{N}$

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

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

Poznámka. Pro $\mathcal{D} = \{0, \ldots, d-1\}, d \geq 2$ 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 $C\colon \Lambda \to \mathcal{D}^*$ prosté zobrazení.
- jednoznačně dekódovatelný, pokud jeho rozšíření $C^*\colon \Lambda^* \to \mathcal{D}^*$,

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

je nesingulární (=prosté), tj. pokud existuje $(C^*)^{-1}\colon \mathcal{D}^* \to \Lambda^*$.

- 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

Kraftova nerovnost

Věta (Kraft, 1949)

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

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

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

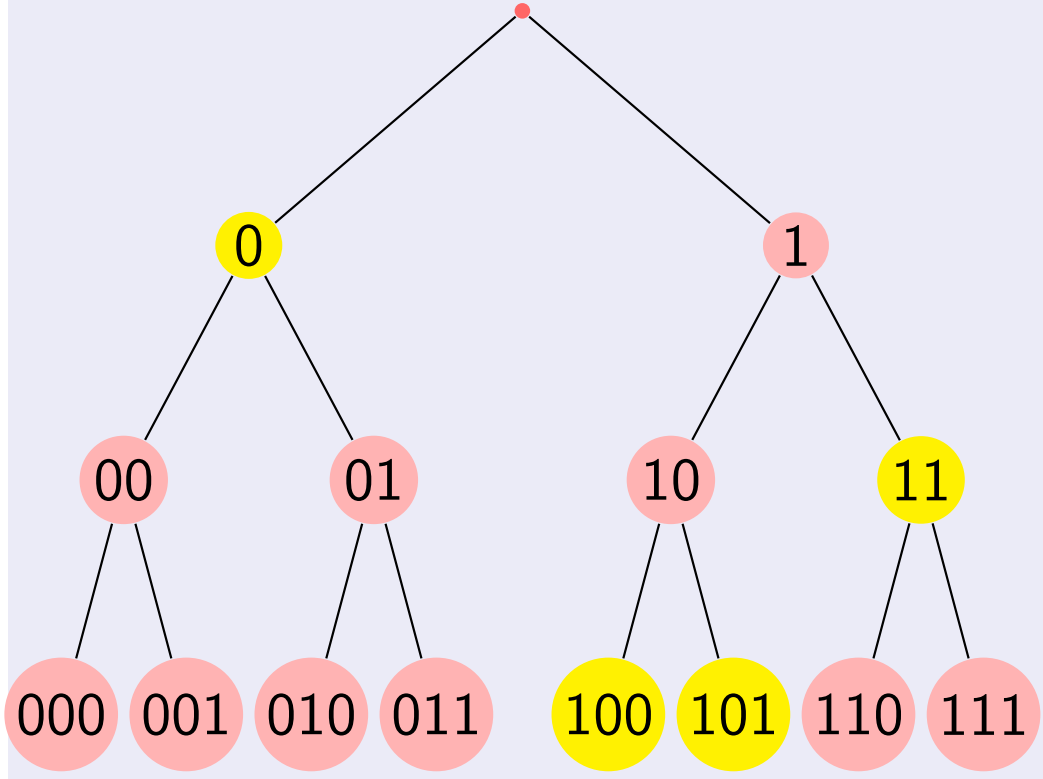
$q_i = d^{-\ell_i}$ je pravděpodobnost vygenerování i-tého kódového slova (délky ℓ_i) náhodným generátorem s rovnoměrným rozdělením; Kraftova nerovnost zní

$$\sum_{i=1}^{m} q_i \leq 1.$$

Důkaz Kraftovy nerovnosti

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 a **poznáme, kdy kódové slovo končí**.

Důkaz Kraftovy nerovnosti (pokr.)

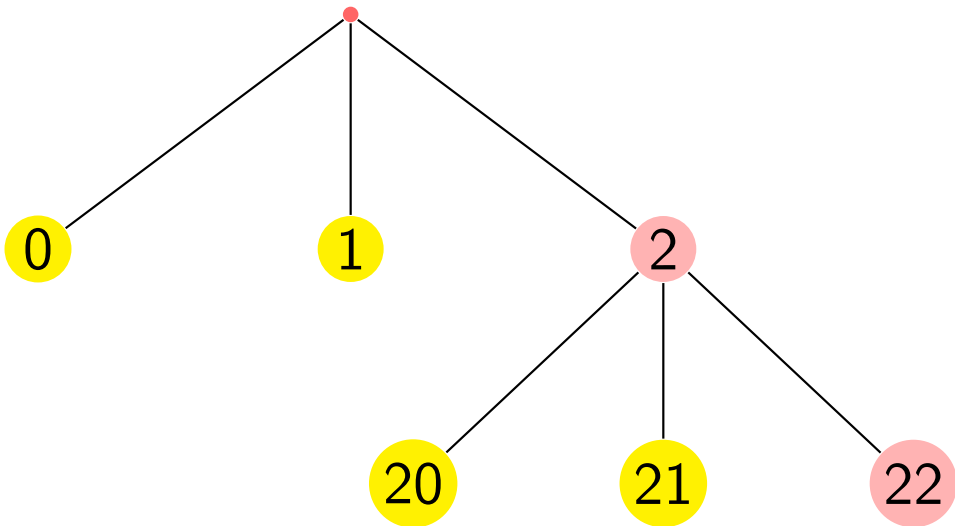
Pokračování důkazu.

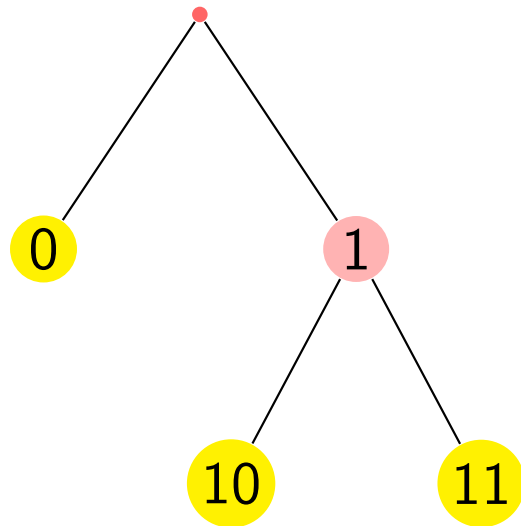
Rovnoměrné generování znaků dá na začátku posloupnosti jediné kódové slovo (s pravděpodobností q_i) nebo "chybu" (neodpovídá žádné kódové slovo); ta má pravděpodobnost $1 - \sum_{i=1}^{m} q_i \geq 0$,

$$\sum_{i=1}^{m} 2^{-\ell_i} = \sum_{i=1}^{m} q_i \leq 1\,.$$

Detekce “chyby”

“Chyba” je možná, právě když Kraftova nerovnost je ostrá; pro binární kód lze vždy docílit rovnost.





Důkaz Kraftovy nerovnosti (pokr.)

Pokračování důkazu.

Obráceně, nechť platí Kraftova nerovnost a $\ell_1 \leq \cdots \leq \ell_m$. 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 $(1 \leq i \leq m-1)$ a jeho následníci vyjmuti
- nalezneme uzel v úrovni ℓ_{i+1}? Je zde

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

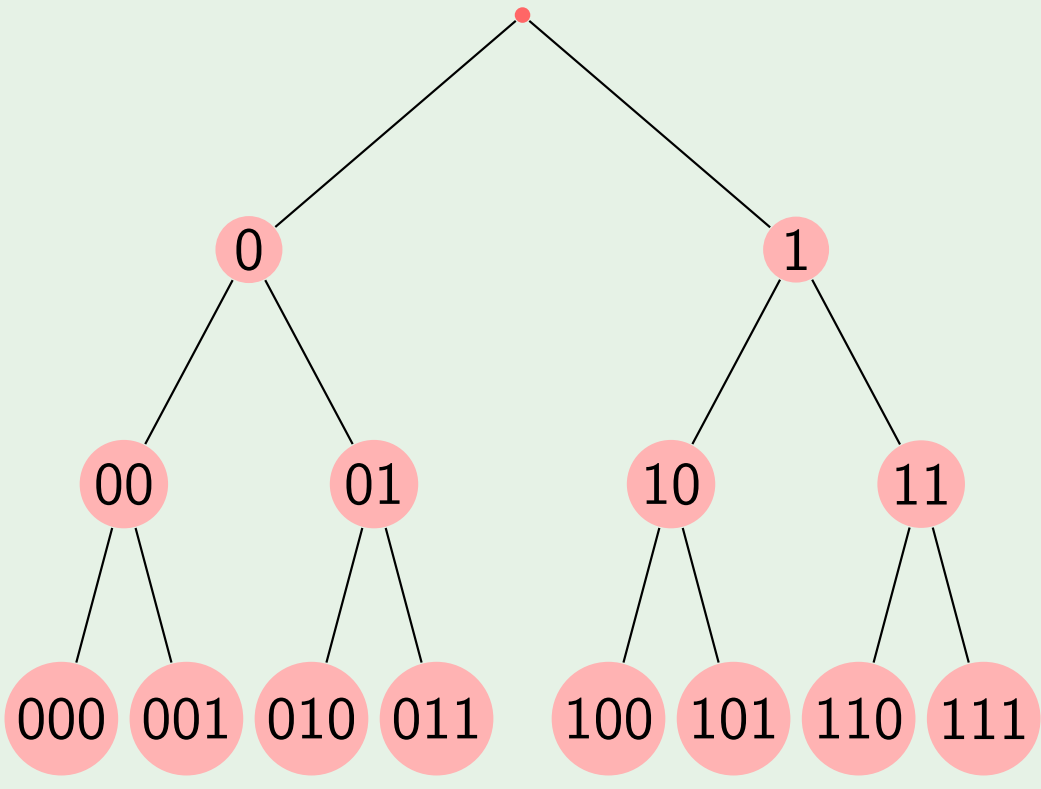
- tento počet je ≥ 1 díky Kraftově nerovnosti (násobme $2^{\ell_{i+1}}$):

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

Ilustrace Kraftovy nerovnosti

Příklad

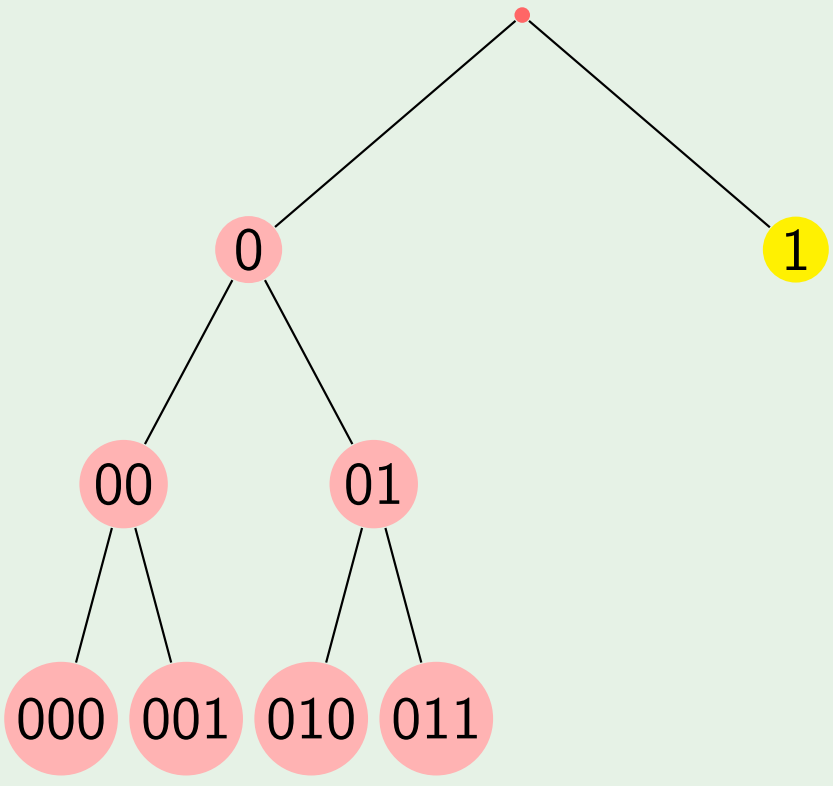
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

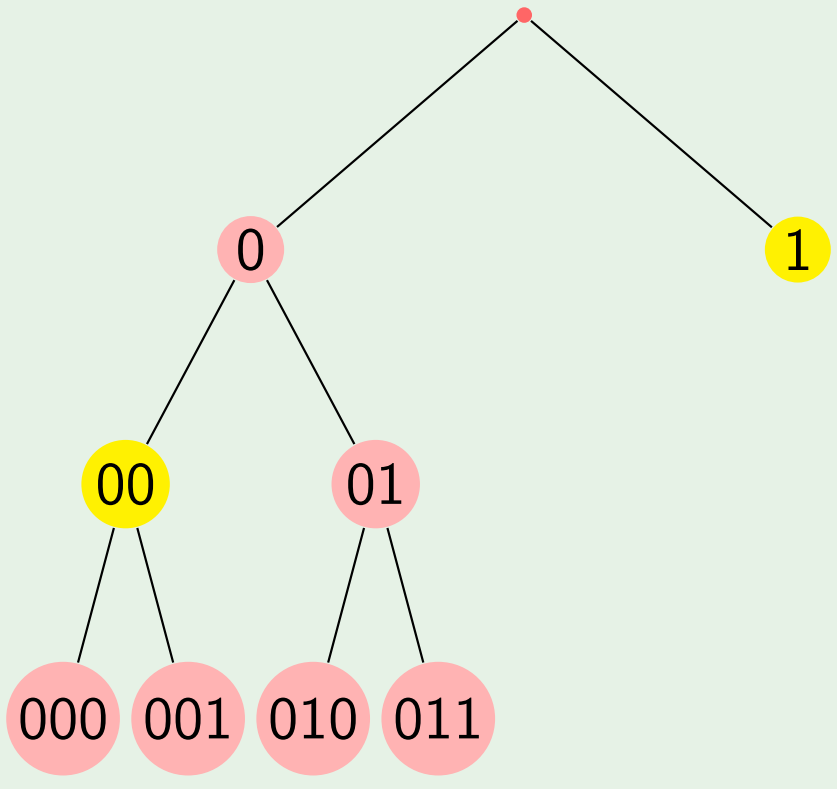
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

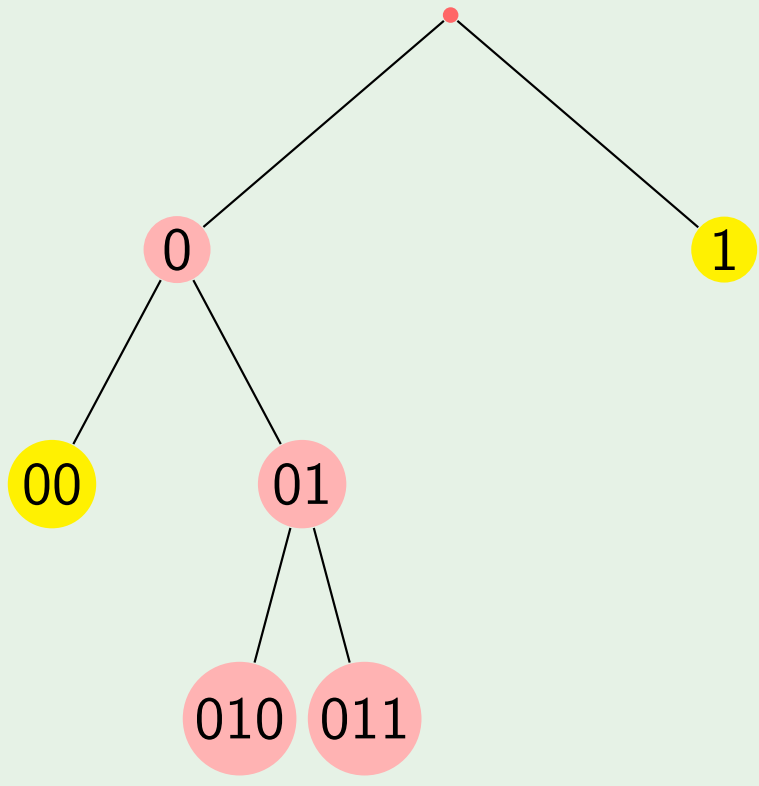
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

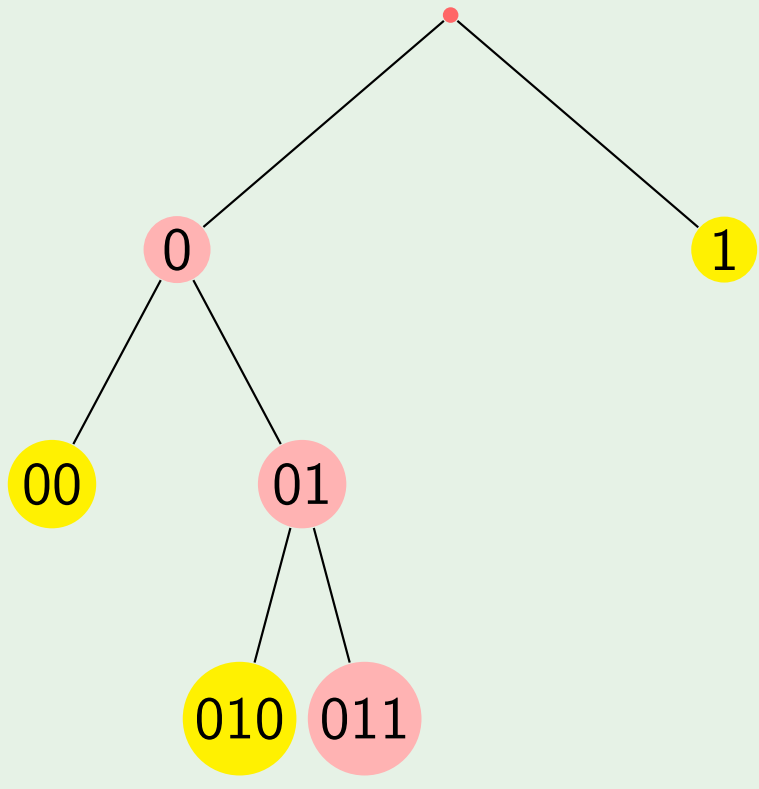
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



Ilustrace Kraftovy nerovnosti

Příklad

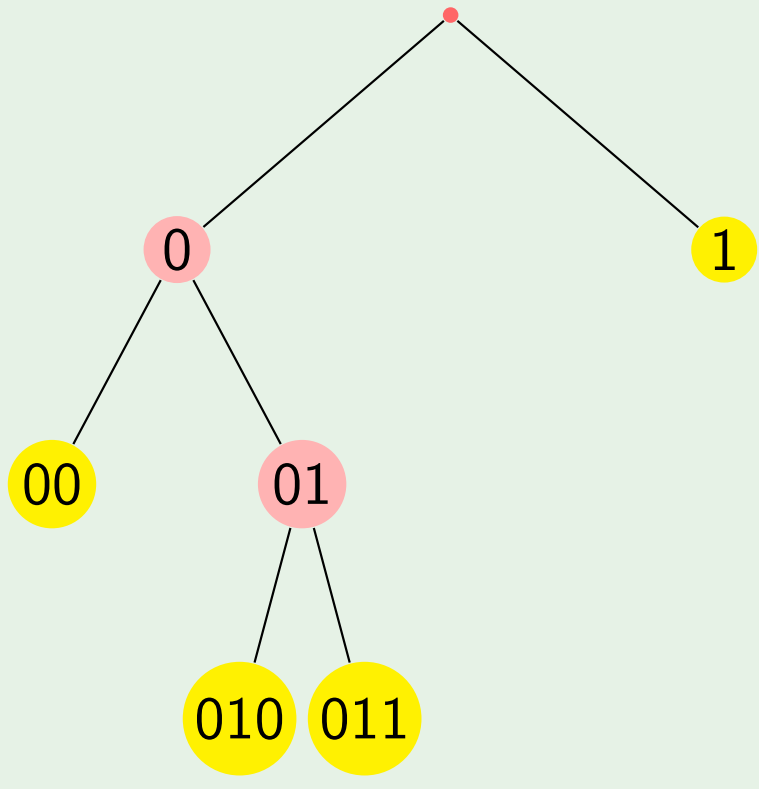
Nalezneme binární kód s délkami kódových slov 1,2,3,3.



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

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

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

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování

Hledáme optimální kód

- Kraftova nerovnost umožňuje konstrukci instantního kódu pomocí zadaných délek kódových slov $\ell_1, \ldots, \ell_m$.
- Pro $d \in \mathbb{N}$ a pravděpodobnostní funkci p_X na m-prvkové množině Λ značíme $p_i = p_X(x_i)$ a hledáme minimum funkce

$$\bar{L}(\ell_1, \ldots, \ell_m) = \sum_{i=1}^{m} p_i \ell_i = -\sum_{i=1}^{m} p_i \log_d q_i$$

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

- Optimální je v Kraftově nerovnosti rovnost a u informační divergence jsme odvodili, že nejlepší jsou $q_i = p_i$, vedoucí na Shannonovu entropii:

$$\bar{L}(\ell_1, \ldots, \ell_m) = -\sum_{i=1}^{m} p_i \log_d p_i = H_d(X) := \frac{H(X)}{\log_2 d}.$$

Hledáme optimální kód (pokr.)

- Optimu se můžeme přiblížit v rámci omezení, že $q_i = d^{-\ell_i}$ nejsou libovolná, neboť $\ell_i \in \mathbb{N}$. Tato úloha je obtížně řešitelná, a proto se prosadily jiné metody.

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
$$L(C) \geq H_d(X) = \frac{H(X)}{\log_2 d}.$$
- Rovnost
$$L(C) = H_d(X)$$
nastává právě tehdy, když $p_X(x) = d^{-\ell(C(x))}$.

Nutné podmínky optimality

Věta

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

- Pokud $p_X(x) > p_X(y)$, pak $\ell(C(x)) \leq \ell(C(y))$.
- Dvě nejméně pravděpodobná kódová slova mají stejnou délku.

Jeli kód navíc *binární*, pak:

- V kódovém stromě pro C má každý list přiřazeno kódové slovo.

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:

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

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

Meze Shannonova kódování

Věta

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

Důkaz.

Snadno plyne z nerovnosti

$$\log_d p_i^{-1} \leq \lceil \log_d p_i^{-1} \rceil < \log_d p_i^{-1} + 1.$$

Poznámka. Střední délka optimálního kódu C^* tedy splňuje

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

Blokové kódování

- uvažujme informační zdroj generující zprávy $(X_1, \ldots, X_n)$ délky n z množiny Λ^n
- binárním **Shannonovým kódem** C_S^n můžeme zakódovat přímo n-tice z Λ^n popsané pomocí $p_{X_1 \ldots X_n}$
- střední délka kódu C_S^n na 1 zdrojový znak $(x_1, \ldots, x_n) \in \Lambda^n$ je

$$L_n(C_S^n) = \frac{L(C_S^n)}{n}$$

- jsou-li $X_1, \ldots, X_n$ nezávislé a stejně rozdělené, pak

$$H(X_1) \leq L_n(C_S^n) < H(X_1) + \frac{1}{n}$$

- lze tedy nalézt kód C_S^n splňující $L_n(C_S^n) - H(X_1) < \varepsilon$ pro libovolné $\varepsilon > 0$

Nevhodné kódování

Věta (Divergence je cena za nevhodné kódování)

Je-li X náhodná veličina s pravděpodobnostní funkcí p a C_S^q je Shannonův kód zkonstruovaný na základě rozdělení q, potom platí:

$$H(X) + D(p\|q) \leq L_p(C_S^q) < H(X) + D(p\|q) + 1.$$

Důkaz.

Odvodíme horní mez, dolní mez se odvodí analogicky:

$$L_p(C_S^q) = \sum_{x\in\Lambda} p(x) \left\lceil \log q^{-1}(x) \right\rceil < \sum_{x\in\Lambda} p(x) \left(\log q^{-1}(x) + 1 \right)$$

$$= \sum_{x\in\Lambda} p(x) \log \left(\frac{p(x)}{q(x)} \frac{1}{p(x)} \right) + p(x) = D(p\|q) + H(X) + 1.$$

- Kódy
- Kraftova nerovnost
- Konstrukce kódů
- Huffmanovo kódování

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.

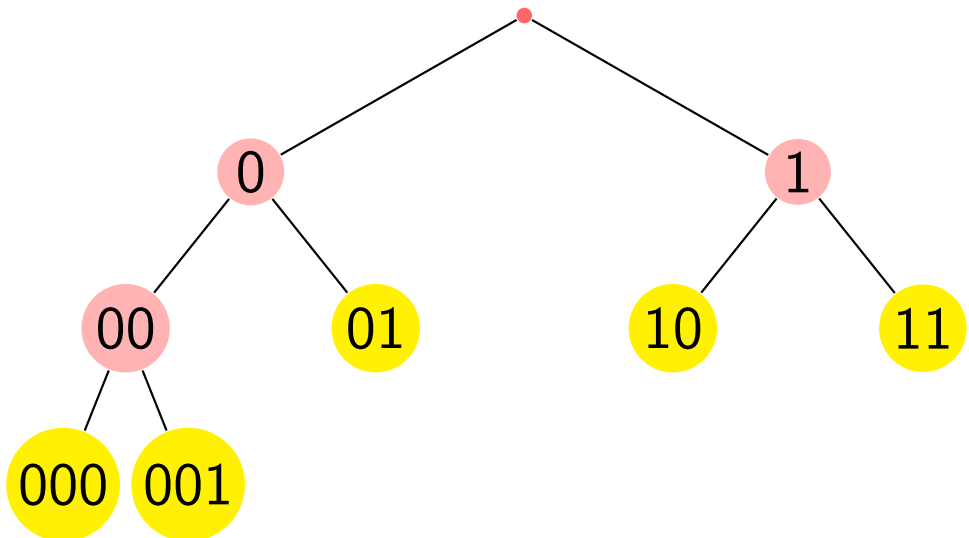
Algoritmus

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



Střední délka Huffmanova kódu $2.3 = 1+$ součet pravděpodobností přidaných do tabulky (vyznačeny červeně).

Huffmanův vs. Shannonův kód

Srovnání délek kódových 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 $L(C_H) \leq L(C)$
- $H(X) \leq L(C_H) < H(X) + 1$

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
- 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.
- Ale kódování **není jednoprůchodové**, je nutno nejdříve načíst celý vstupní soubor.
- Umíme optimálně komprimovat zdroje s pamětí?

Part IV

Bezšumový přenosový kanál s pamětí

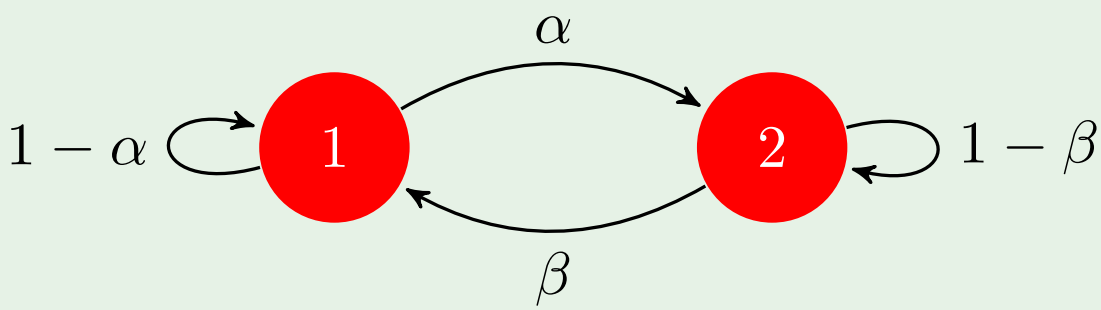
Zdroj s pamětí

- 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 určíme pomocí **rychlosti entropie**.

Příklad: 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.

Motivace

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

- Zdroj $(X_1, \ldots, X_n)$ má sdruženou entropii $H(X_1, \ldots, X_n)$.
- Pro $n \to \infty$ není entropie zdroje $(X_1, \ldots, X_n)$ shora omezena, neboť pro $|\Lambda| \geq 2$ platí

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

Řešení

Hledejme entropii na jeden znak zprávy!

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

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

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 $H(X_2|X_1),\ H(X_3|X_2,X_1),\dots$ nerostoucí a její limita $\widetilde{H}((X_n)_{n\in\mathbb{N}})$ existuje.

Důkaz.

Pro $n \geq 2$ platí

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

kde $\leq$ 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.

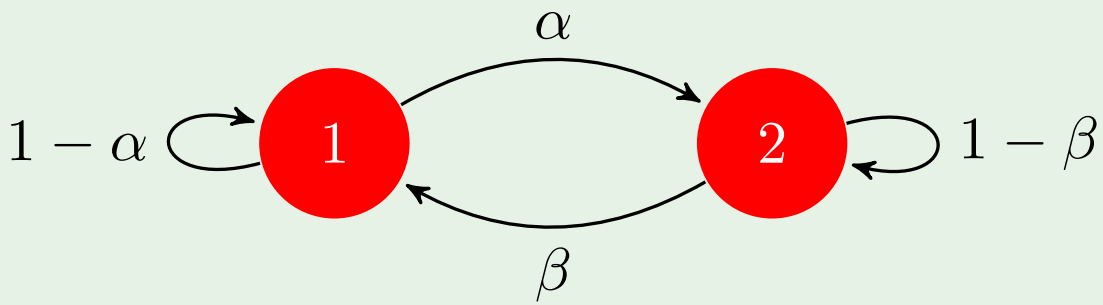
Markovský zdroj

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

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

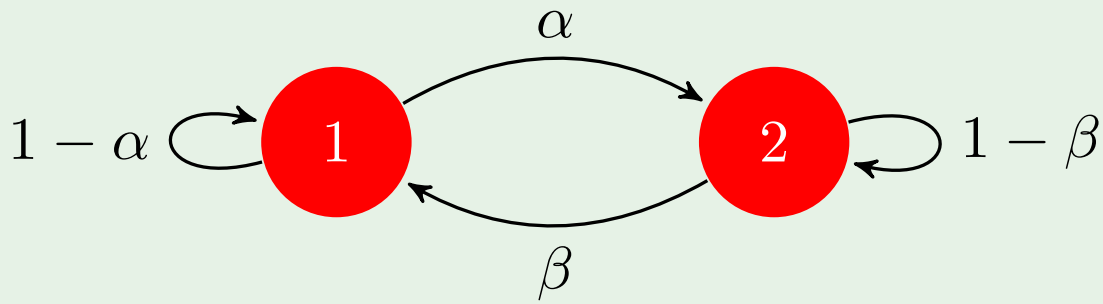
Důkaz.

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

Příklad markovského zdroje

Příklad (pokr.)

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

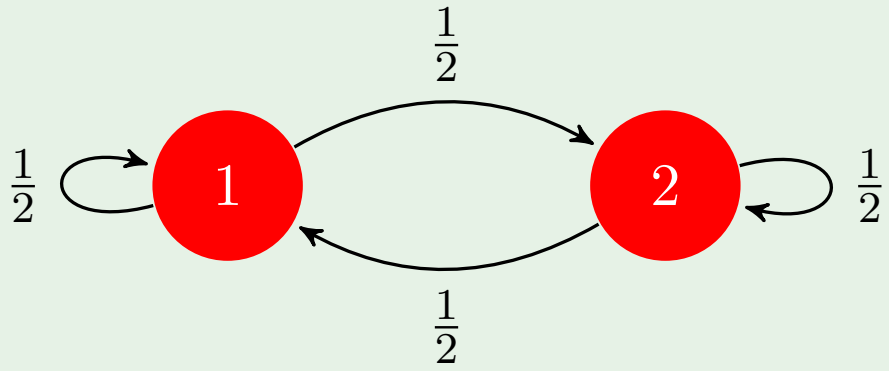


Počáteční rozdělení $\mathbf{p}(0) = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$. Rychlost entropie je

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

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

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



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

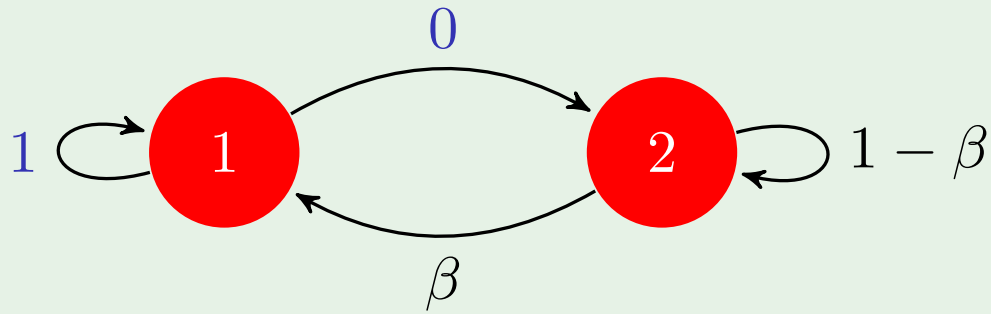
Jde o **bezpaměťový** zdroj informace a ve zprávě $x_1 x_2 \ldots$ má každý řetězec

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

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

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

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



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

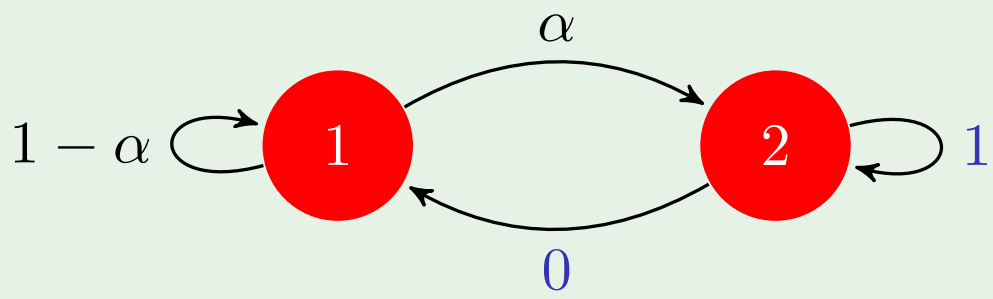
Jde o **deterministický** zdroj informace: zpráva

$$111\ldots$$

vznikne s pravděpodobností 1.

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

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



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

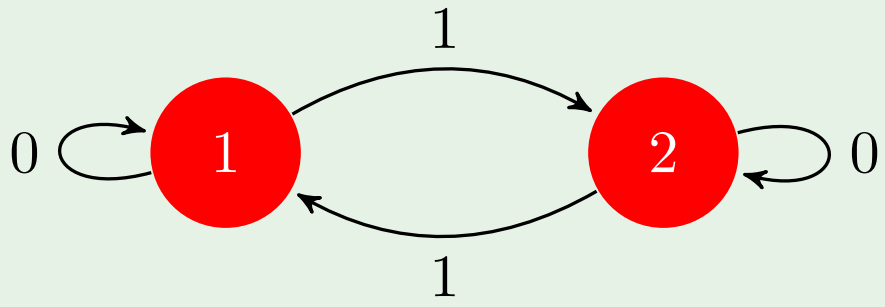
Jde o **deterministický** zdroj informace: zpráva

$$222\ldots$$

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 $1/2$.

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.

Part V

Přenos kanálem se šumem

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

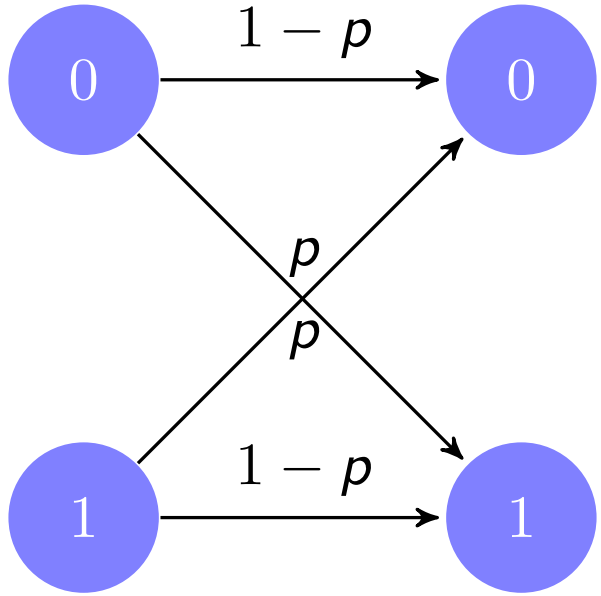
- Λ je m-prvková vstupní abeceda
- Ω je n-prvková výstupní abeceda
- $\mathbf{P}$ je matice $m \times n$ podmíněných pravděpodobností:

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

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 uzavřené omezené množině Δ_m, suprema nabývá
- 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:

- $C(\mathcal{K}) \geq 0$
- $C(\mathcal{K}) \leq \log |\Lambda|$
- $C(\mathcal{K}) \leq \log |\Omega|$

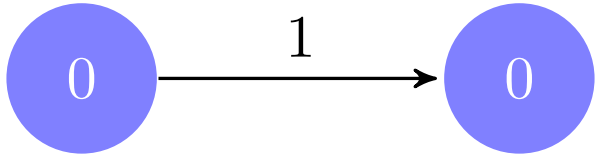
Výpočet kapacity

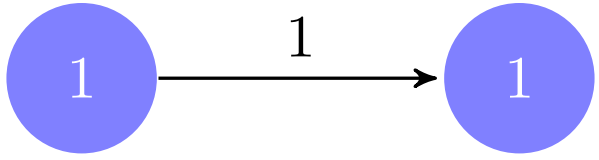
- hledáme maximum rozdílu $H(Y) - H(Y|X)$ pro X
- výpočet je snadný pro kanály, v nichž se pravděpodobnostní funkce $p_{Y|X}(\cdot|x_i)$, $p_{Y|X}(\cdot|x_j)$ liší pouze permutací odpovídajících pravděpodobností; pak
- matice přechodu $\mathbf{P}$ je tvořena permutacemi jednoho řádku
- $H(Y|X = x_i)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

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

Binární bezšumový kanál

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

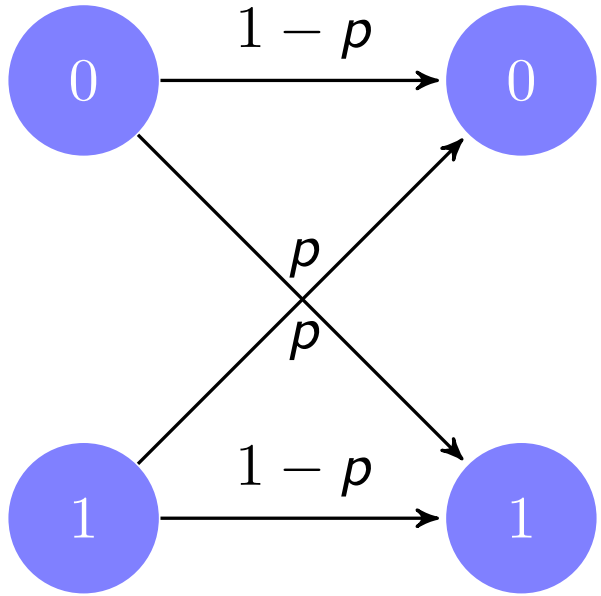




Kapacita $C(\mathcal{K}) = 1$

Binární symetrický kanál

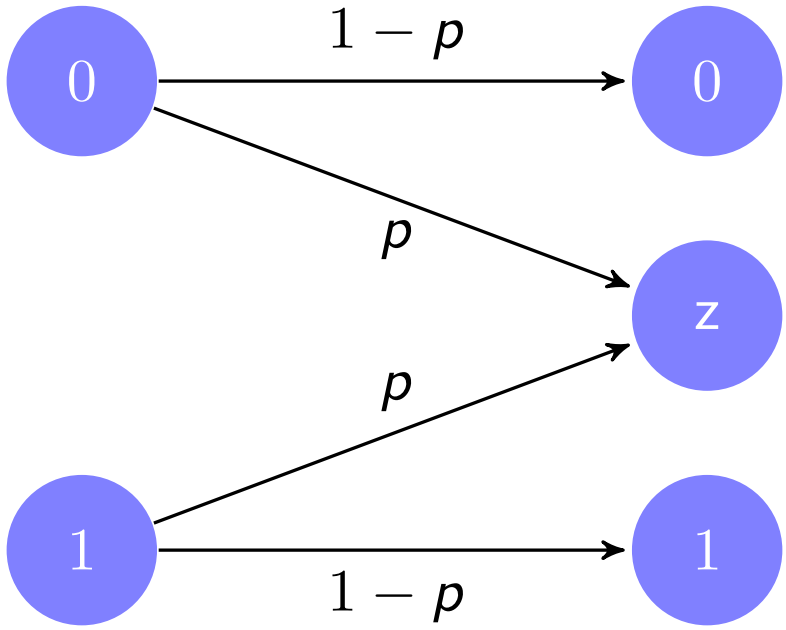
$\Lambda = \Omega = \{0, 1\}$



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

Binární kanál se zámlkou

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



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

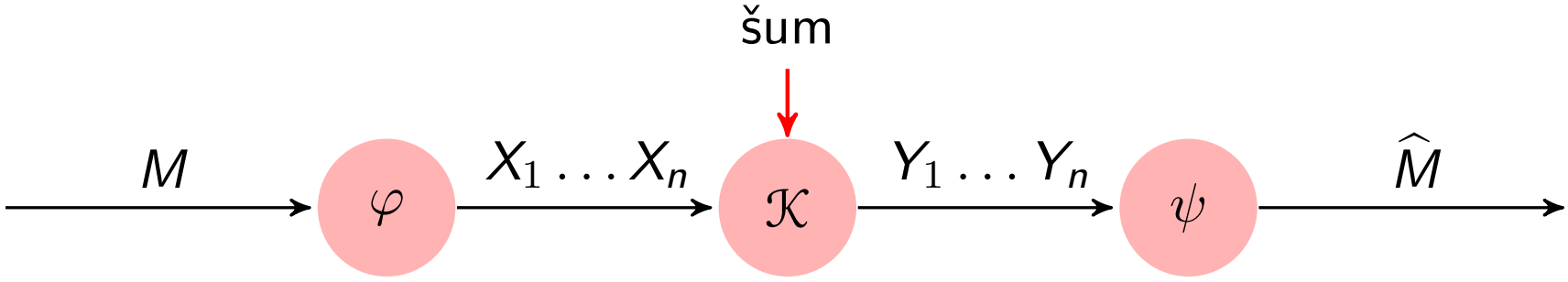
Zámlka lepší než záměna

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

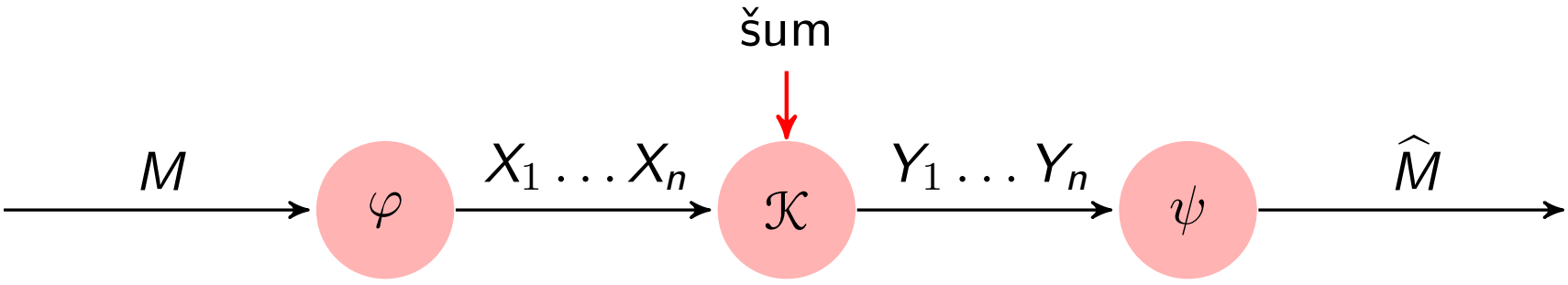
1. **zdroj** je náhodná veličina M s rovnoměrným rozdělením na konečné množině zpráv $\mathcal{M}$
2. **kodér** je zobrazení
$$\varphi : \mathcal{M} \to \Lambda^n,$$
které každé zdrojové zprávě $i \in \mathcal{M}$ přiřazuje vstupní kódové slovo $\varphi(i) = x_1 \dots x_n \in \Lambda^n$
3. **bezpaměťový 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

$$\lambda_i = P\left[\psi(Y_1 \dots Y_n) \neq i \mid X_1 \dots X_n = \varphi(i)\right]$$

- 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
- pro praktické použití se tedy nehodí...
- honba za kódy, jejichž rychlost se přibližuje kapacitě kanálu, začala v 50.letech minulého století a pokračuje dodnes

Part VI

Dodatek

Jensenova nerovnost

Věta

Nechť $f : (a, b) \to \mathbb{R}$ 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

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

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

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

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.