# Elementy teorii informacji i kodowania

## Entropia, nierówność Krafta, kodowanie optymalne

Marcin Jenczmyk
m.jenczmyk@knm.katowice.pl

17 kwietnia 2015

# Elementy teorii informacji

Niech $S = \{x_1, \ldots, x_q\}$ oznacza alfabet, zbiór zdarzeń, a $P = \{p_1, \ldots, p_q\}$ będzie stowarzyszonym z nim zbiorem prawdopodobieństw takim, że $P(x_i) = p_i$ oraz $\sum_{i=1}^{q} p_i = 1$.

# Elementy teorii informacji

Niech $S = \{x_1, \ldots, x_q\}$ oznacza alfabet, zbiór zdarzeń, a $P = \{p_1, \ldots, p_q\}$ będzie stowarzyszonym z nim zbiorem prawdopodobieństw takim, że $P(x_i) = p_i$ oraz $\sum_{i=1}^{q} p_i = 1$. Ponadto załóżmy, że źródło $S$ jest *źródłem bez pamięci*, tzn. że kolejne znaki są generowane niezależnie od siebie.

# Elementy teorii informacji

## Informacja

Ilością informacji stowarzyszonej z zdarzeniem nazywamy funkcję $I$ określoną na zbiorze zdarzeń, zadaną wzorem

$$I(x) = -\log_k P(x)$$

dla pewnego $k > 0$ oraz $k \neq 1$.

# Elementy teorii informacji

## Informacja

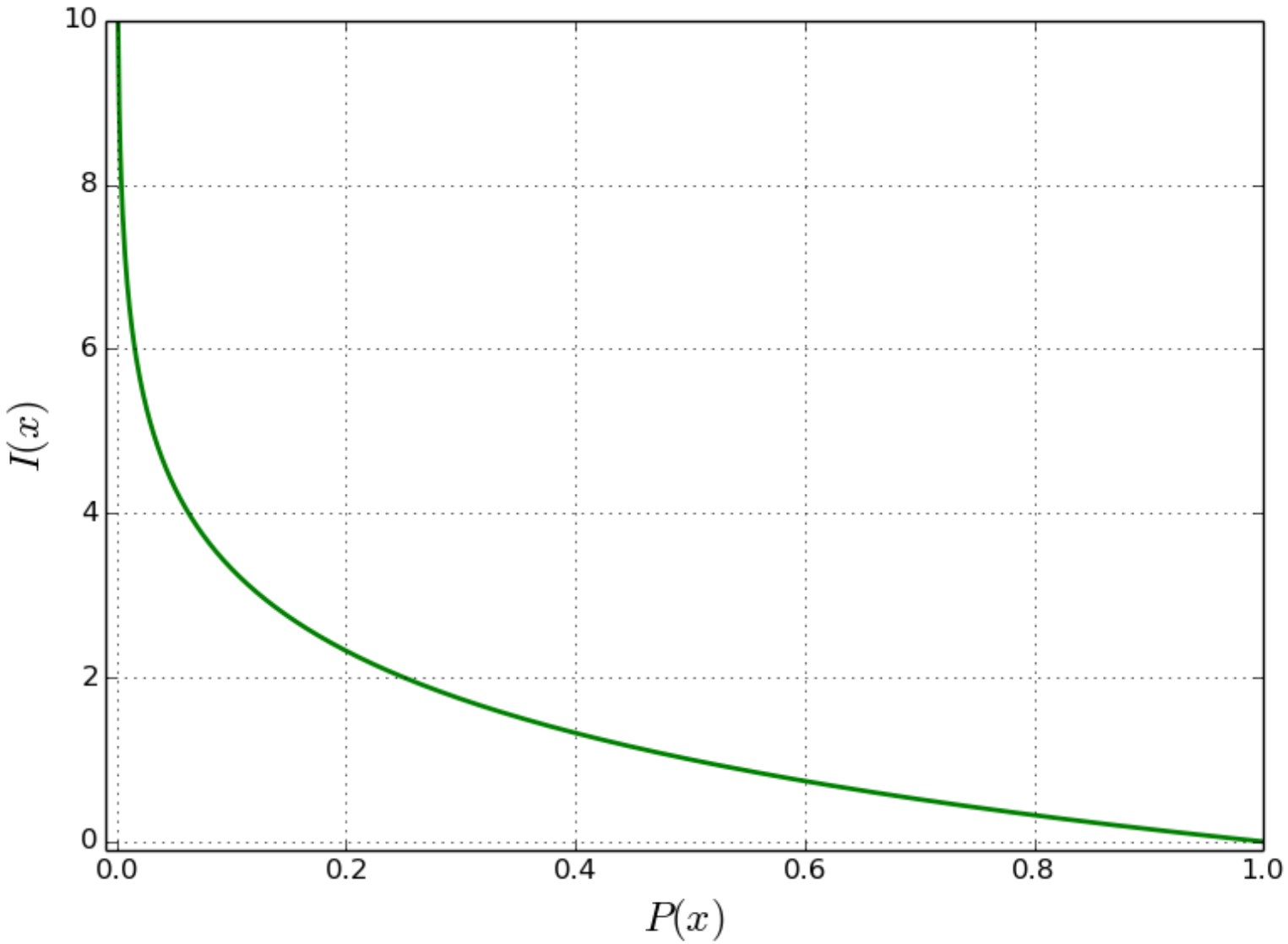
Ilością informacji stowarzyszonej z zdarzeniem nazywamy funkcję $I$ określoną na zbiorze zdarzeń, zadaną wzorem

$$I(x) = -\log_k P(x)$$

dla pewnego $k > 0$ oraz $k \neq 1$.

W zależności od podstawy logarytmu ilość informacji może być mierzona w różnych jednostkach. Dla $k = 2$ jednostką informacji jest bit, dla $k = e$ nat, a dla $k = 10$ Hartley.

# Elementy teorii informacji



# Elementy teorii informacji

## Entropia

Entropią stowarzyszoną z źródłem $S = \{x_1, \ldots, x_q\}$ nazywamy funkcję

$$H(S) = H(p_1, \ldots, p_q) = \sum_{i=1}^{q} P(x_i) \log P(x_i) = \sum_{i=1}^{q} P(x_i) I(x_i).$$

Funkcja ta jest wartością oczekiwaną ilości informacji ujawnianej przez źródło.

# Elementy teorii informacji

## Przykład

Rozważmy źródło $S = \{x_1, x_2, x_3\}$ takie, że $P(x_1) = \frac{1}{2}$, a $P(x_2) = P(x_3) = \frac{1}{4}$. Wówczas

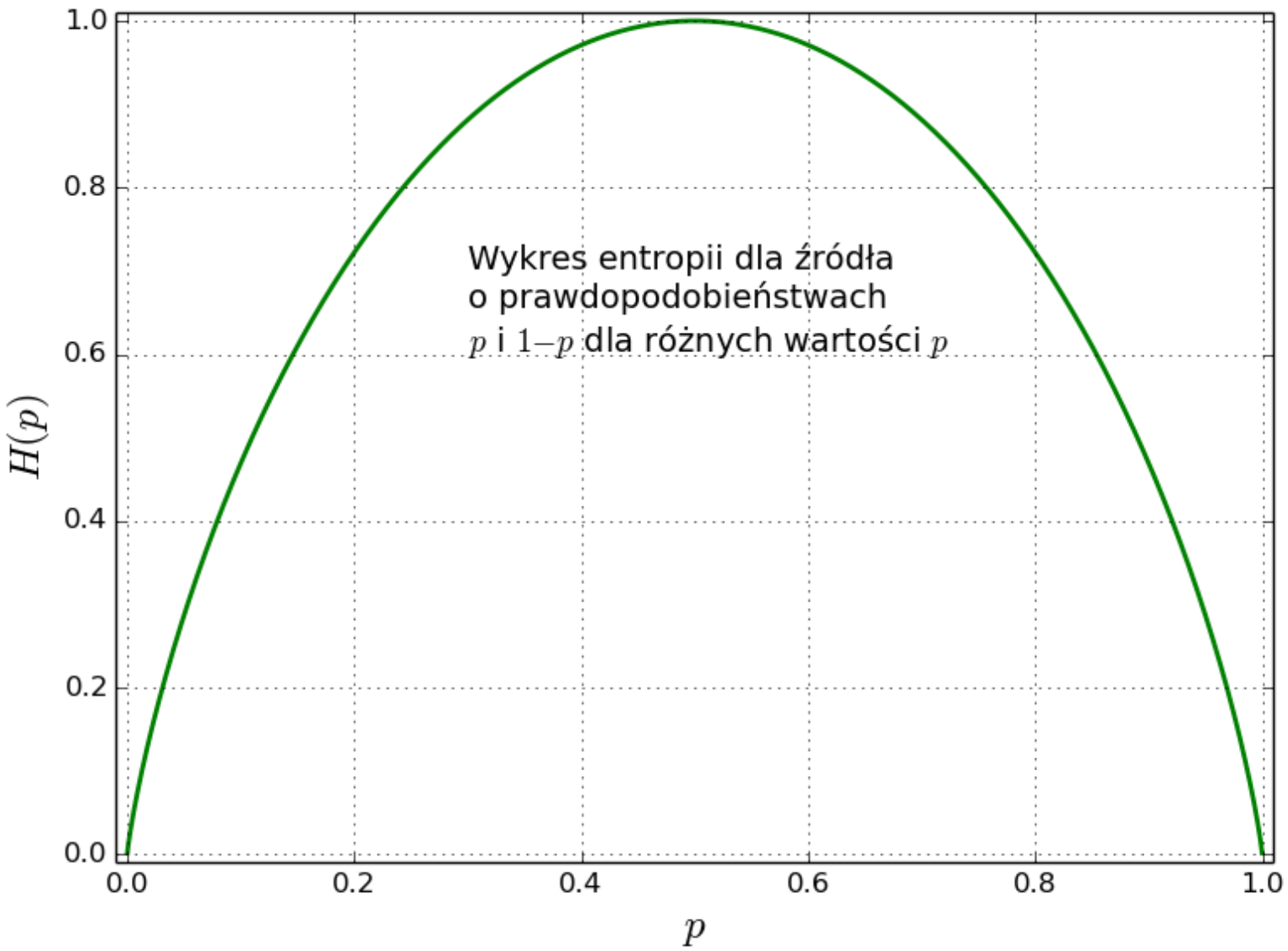
$$I(x_1) = -\log \frac{1}{2} = \log 2 = 1,$$

$$I(x_2) = I(x_3) = -\log \frac{1}{4} = \log 4 = 2,$$

a entropia źródła wynosi

$$H(S) = \frac{1}{2} I(x_1) + \frac{1}{4} I(x_2) + \frac{1}{4} I(x_3) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}.$$

# Elementy teorii informacji



# Elementy teorii informacji

## Twierdzenie

*Dla dowolnego zbioru prawdopodobieństw $p_1, \ldots, p_q$ zachodzi oszacowanie*

$$0 = H(1, 0, \ldots, 0) \leqslant H(p_1, \ldots, p_q) \leqslant H\left(\frac{1}{q}, \ldots, \frac{1}{q}\right) = \log q.$$

# Elementy teorii informacji

## Dowód.

Ponieważ $\sum_{i=1}^{q} p_i = 1$, to $p_q = 1 - \sum_{i=1}^{q-1} p_i$ oraz

$$\begin{aligned} H(p_1, \ldots, p_q) &= H(p_1, \ldots, p_{q-1}) \\ &= -\sum_{i=1}^{q-1} p_i \log p_i - \left(1 - \sum_{i=1}^{q-1} p_i\right) \log \left(1 - \sum_{i=1}^{q-1} p_i\right). \end{aligned}$$

# Elementy teorii informacji

## Dowód.

Ponieważ $\sum_{i=1}^{q} p_i = 1$, to $p_q = 1 - \sum_{i=1}^{q-1} p_i$ oraz

$$\begin{aligned} H(p_1, \ldots, p_q) &= H(p_1, \ldots, p_{q-1}) \\ &= -\sum_{i=1}^{q-1} p_i \log p_i - \left(1 - \sum_{i=1}^{q-1} p_i\right) \log \left(1 - \sum_{i=1}^{q-1} p_i\right). \end{aligned}$$

Funkcja może mieć ekstremum w punkcie, gdy

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= -(\log p_i + \log e) + \log \left(1 - \sum_{i=1}^{q-1} p_i\right) + \log e \\ &= -\log p_i + \log \left(1 - \sum_{i=1}^{q-1} p_i\right) = \log \frac{p_q}{p_i} = 0, \end{aligned}$$

# Elementy teorii informacji

## Dowód (cd.)

Tak się dzieje gdy

$$\log \frac{p_q}{p_i} = 0 \Leftrightarrow \frac{p_q}{p_i} = 1 \Leftrightarrow p_i = p_q = \frac{1}{q}.$$

# Elementy teorii informacji

## Dowód (cd.)

Tak się dzieje gdy

$$\log \frac{p_q}{p_i} = 0 \Leftrightarrow \frac{p_q}{p_i} = 1 \Leftrightarrow p_i = p_q = \frac{1}{q}.$$

W badanym punkcie

$$\frac{\partial^2 H}{\partial p_i \partial p_i} = -\frac{\log e(p_i + p_q)}{p_q p_i} = -\log e \left( \frac{1}{p_i} + \frac{1}{p_q} \right) = -2q \log e$$

# Elementy teorii informacji

## Dowód (cd.)

Tak się dzieje gdy

$$\log \frac{p_q}{p_i} = 0 \Leftrightarrow \frac{p_q}{p_i} = 1 \Leftrightarrow p_i = p_q = \frac{1}{q}.$$

W badanym punkcie

$$\frac{\partial^2 H}{\partial p_i \partial p_i} = -\frac{\log e(p_i + p_q)}{p_q p_i} = -\log e \left( \frac{1}{p_i} + \frac{1}{p_q} \right) = -2q \log e$$

oraz

$$\frac{\partial^2 H}{\partial p_i \partial p_j} = -\frac{\log e}{p_q} = -q \log e.$$

# Elementy teorii informacji

## Dowód (cd.)

Macierz drugich pochodnych w badanym punkcie jest postaci

$$\begin{bmatrix} -2q\log e & -q\log e & \dots & -q\log e \\ -q\log e & -2q\log e & \dots & -q\log e \\ \vdots & \vdots & \dots & \vdots \\ -q\log e & -q\log e & \dots & -q\log e \end{bmatrix}$$

# Elementy teorii informacji

## Dowód (cd.)

Macierz drugich pochodnych w badanym punkcie jest postaci

$$\begin{bmatrix} -2q\log e & -q\log e & \dots & -q\log e \\ -q\log e & -2q\log e & \dots & -q\log e \\ \vdots & \vdots & \dots & \vdots \\ -q\log e & -q\log e & \dots & -q\log e \end{bmatrix}$$

Macierz ta jest ujemnie określona, więc punkt jest maksimum entropii.

# Elementy kodowania

## Kodowanie

Kodowaniem elementów alfabetu $S = \{x_1, \ldots, x_q\}$ o prawdopodobieństwach $P = \{p_1, \ldots, p_q\}$ przy użyciu alfabetu $C = \{c_1, \ldots, c_n\}$ nazywamy proces przyporządkowania elementom $S$ elementów lub ciągów znaków (słów kodowych) alfabetu $C$.

# Elementy kodowania

## Kodowanie

Kodowaniem elementów alfabetu $S = \{x_1, \ldots, x_q\}$ o prawdopodobieństwach $P = \{p_1, \ldots, p_q\}$ przy użyciu alfabetu $C = \{c_1, \ldots, c_n\}$ nazywamy proces przyporządkowania elementom $S$ elementów lub ciągów znaków (słów kodowych) alfabetu $C$.

## Przykład (kod binarny)

| Elementy alfabetu | Słowa kodowe |
|---|---|
| $s_1$ | 00 |
| $s_2$ | 01 |
| $s_3$ | 10 |
| $s_4$ | 11 |

# EElementy kodowania

## Kod jednoznacznie dekodowalny

Kod nazywamy jednoznacznie dekodowalnym, jeśli istnieje tylko jeden sposób podziału ciągu słów kodu $c_{i_1}, c_{i_2}, \ldots, c_{i_k}$ na oddzielne słowa kodu, tzn. jeśli

$$c_{i_1}, c_{i_2}, \ldots, c_{i_k} = c_{j_1}, c_{j_2}, \ldots, c_{j_k}$$

to $l = 1, \ldots, k$ zachodzi

$$c_{i_l} = c_{j_l}.$$

# EElementy kodowania

## Kod jednoznacznie dekodowalny

Kod nazywamy jednoznacznie dekodowalnym, jeśli istnieje tylko jeden sposób podziału ciągu słów kodu $c_{i_1}, c_{i_2}, \ldots, c_{i_k}$ na oddzielne słowa kodu, tzn. jeśli

$$c_{i_1}, c_{i_2}, \ldots, c_{i_k} = c_{j_1}, c_{j_2}, \ldots, c_{j_k}$$

to $l = 1, \ldots, k$ zachodzi

$$c_{i_l} = c_{j_l}.$$

## Kod prefixowy

Kod nazywamy kodem prefixowym, jeśli żadne słowo kodowe nie jest przedrostkiem innego słowa kodowego.

# Elementy kodowania

## Przykłady

| Znak | $K_1$ | $K_2$ | $K_3$ | $K_4$ |
|---|---|---|---|---|
| A | 0 | 0 | 00 | 00 |
| B | 1 | 11 | 01 | 01 |
| C | 01 | 01 | 10 | 1 |

# Elementy kodowania

## Nierówność Krafta

Binarny kod przedrostkowy złożony ze słów kodu $\{c_1, \ldots, c_n\}$ o długościach $\{l_1, \ldots, l_n\}$ istnieje $\Leftrightarrow$

$$\sum_{i=1}^{n} 2^{-l_i} \leqslant 1. \tag{1}$$

# Elementy kodowania

## Nierówność Krafta

Binarny kod przedrostkowy złożony ze słów kodu $\{c_1, \ldots, c_n\}$ o długościach $\{l_1, \ldots, l_n\}$ istnieje $\Leftrightarrow$

$$\sum_{i=1}^{n} 2^{-l_i} \leqslant 1. \tag{1}$$

## Dowód.

($\Rightarrow$) Mając binarny kod przedrostkowy umieśćmy słowa kodu w drzewie binarnym zgodnie z kodami słów i uzupełnijmy je do drzewa pełnego, o $2^{l_{\max}}$ liściach. $i$-ty wierzchołek "zajmuje" $2^{l_{\max}-l_i}$ liści drzewa. Wobec tego musi zachodzić

$$\sum_{i=1}^{n} 2^{l_{\max}-l_i} \leqslant 2^{l_{\max}} \Leftrightarrow \sum_{i=1}^{n} 2^{-l_i} \leqslant 1.$$

# Elementy kodowania

## Dowód.

($\Leftarrow$) Przeciwnie, załóżmy, że (1) zachodzi. Wówczas możemy umieścić słowa w wybranych węzłach, tworząc w ten sposób kody słów. Zauważmy, że liści nie może zabraknąć - po umieszczeniu w drzewie $j$-tego wierzchołka ilość "zablokowanych" to

$$\sum_{i=1}^{j} 2^{l_{\max}-l_i} \leqslant 2^{l_{\max}} \sum_{i=1}^{j} 2^{-l_i} 2^{l_{\max}} \sum_{i=1}^{n} 2^{-l_i} \leqslant 2^{l_{\max}}.$$

□

# Elementy kodowania

## Twierdzenie

*Dla dowolnego binarnego kodu przedrostkowego dla którego średnia długość słowa kodowego* $L_{śr} = \sum p_i l_i$ *zachodzi nierówność*

$$L_{śr} \geqslant H(S)$$

*Ponadto istnieje binarny kod przedrostkowy dla którego*

$$L_{śr} < H(S) + 1$$

# Elementy kodowania

## Dowód.

Na początku udowodnimy część pierwszą tezy.

$$H(S) - L_{\text{śr}} = -\sum p_i \log p_i - \sum p_i l_i$$

$$= -\sum p_i \left(\log p_i + \log 2^{l_i}\right) = -\sum (p_i \log p_i 2^{l_i}) = \sum pi \log \frac{1}{p_i 2^{l_i}}$$

$$\leqslant \sum p_i \left(\frac{1}{p_i 2^{l_i}} - 1\right) \log e = \log e \sum (2^{-l_i} - p_i) \leqslant 0$$

Dowód drugiej części. Niech $l_i = \lceil -\log p_i \rceil$.

$$\sum 2^{-l_i} \leqslant \sum p_i = 1$$

Ponadto

$$l_i < -\log p_i + 1 \Leftrightarrow p_i l_i < -p_i \log p_i + p_i \Leftrightarrow L_{\text{śr}} < H(S) + 1$$

# Bibliografia


- Adam Drozdek, *Wprowadzenie do kompresji danych*, WNT, Warszawa 1999
- Norman Abromson, *Teoria informacji i kodowania*, PWN, 1969

# Dziękuję za (nie)uwagę!