# Kodowanie, Kompresja, Kryptografia

**konspekt I, 7.III.2023**

## Literatura


1. T. K. Carne, Codes and Cryptography, skrypt Cambridge University, 2015
   https://www.dpmms.cam.ac.uk/~tkc/CodesandCryptography/CodesandCryptography.pdf
2. J.I. Hall, Notes on Coding Theory, Michigan State University, 2015
   https://users.math.msu.edu/users/halljo/classes/codenotes/coding-notes.html
3. M. Cover, J.A. Thomas, Elements of information theory, Wiley 2006
   https://www.wiley.com/en-us/Elements+of+Information+Theory%2C+2nd+Edition-p-9780471241959
4. D.R. Stinson, M.B. Paterson, Cryptography: Theory and Practice,
   Taylor & Francis 2019; po polsku: Kryptografia w teorii i praktyce, PWN 2021
   https://ksiegarnia.pwn.pl/Kryptografia-W-teorii-i-praktyce,897655934,p.html


Kodowanie, kompresja, kryptografia — pojęcia związane z przechowywaniem i przesyłaniem informacji.

Ocena *ilości* informacji w kodowanej sekwencji

— stratna transmisja kodu, ocena ilości utraconej informacji, niezbędna redundancja dla bezstratnego przesyłu

— kompresja, ocena redundancji, skrócenie zapisu bez utraty informacji

— szyfrowanie, rozpraszanie statystycznych korealcji komunikatu, tak aby proces łamania szyfru był maksymalnie utrudniony

$\mathcal{A}$ — alfabet wejściowy, np. $\{a, b, ..., z, A, B, ..., Z, ...\}$
$\mathcal{B}$ — alfabet kodu, transmisji, np. $\{0, 1\}$
$c : \mathcal{A} \to \mathcal{B}^*$ — funkcja kodująca, np. ASCII, $c(a) = 01100001$
$\kappa : \mathcal{B}^* \to \mathcal{F}$ — processing (kompresja, sumy kontrolne...) przed transmisją.
$\kappa^{-1} : \mathcal{F} \dashrightarrow \mathcal{B}^*$ — odwrotny processing (częściowy) po transmisji
$c^{-1} : \mathcal{B}^* \to \mathcal{A}$ — dekodowanie

## Prawdopodobieństwo

$\mathcal{E}$ — zbiór zdarzeń elementarnych (tu: skończony)
$\Omega$ — $\sigma$-algebra zdarzeń, $A \in \Omega \Leftrightarrow A \subset \mathcal{E}$

1. $\emptyset \in \Omega$
2. Jeśli $A \in \Omega$, to także $A^c = \mathcal{E} - A \in \Omega$
3. Jeśli $A_1, A_2, ... \in \Omega$, to także $\bigcup_{i=1}^{\infty} A_i \in \Omega$.

Ponieważ $(A^c \cup B^c)^c = A \cap B$, jeśli $A, B \in \Omega$, wtedy także $A \cap B \in \Omega$.

Tutaj, ponieważ $\mathcal{E}$ jest zbiorem skończonym, bierzemy $\Omega = \mathcal{P}(\mathcal{E})$.

Rozkład prawdopodobnieństwa na $\Omega$: $P : \Omega \to [0, 1]$

1. $P(\emptyset) = 0$
2. $P(\mathcal{E}) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ jeśli $A_i \cap A_j = \emptyset \quad \forall\, i, j.$

Niektóre własności

1. $P(A^c) = 1 - P(A)$
2. Jeśli $A \subset B$, to $P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. Dla skończonej przestrzeni $\mathcal{E} = \{a_1, a_2, \ldots a_n\}$ rozkład $P$ na $\Omega$ jest w całości wyznaczony przez liczby $p_i = P(\{a_i\})$, $i = 1, 2, ..., n$.

## Zmienna losowa i jej rozkład

Funkcja mierzalna $X : \mathcal{E} \to \mathbb{R}$, to jest taka że $\forall\, (a, b) \quad X^{-1}(a, b) \in \Omega$.

Rozkład zmiennej losowej $X$

$$P_X(a < X < b) = P(X^{-1}(a, b)) \,.$$

**Statystyka**

Wartość oczekiwana zmiennej losowej: $\mathbb{E}[X] = \int x\, P_X(X = x)\, dx$

Wariancja $X$: $\mathrm{Var}\,(X) = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big] = \int (x - \mathbb{E}[X])^2 P_X(X = x)\, dx$

$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \qquad \sigma(X) = \sqrt{\mathrm{Var}\,(X)}$

Kowariancja: $\mathrm{Cov}\,(X, Y) = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big]$

Korelacja: $r_{XY} = \dfrac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$

W naszym przypadku, gdy $\mathcal{E} = \{a_1, a_2, ..., a_n\}, \quad P(a_i) = p_i, \quad X(a_i) = x_i$

$$\mathbb{E}[X] = \sum_{i=1}^{n} p_i x_i\,, \qquad \mathrm{Var}\,(X) = \sum_{i=1}^{n} p_i (x_i - \mathbb{E}(X))^2\,, \qquad \text{itd.}$$

**Prawdopodobieństwo warunkowe**

Dla $A, B \in \Omega$, jeśli $P(B) \neq 0$, określamy $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Przykład: W grupie 100-osobowej jest 55 kobiet i 45 mężczyzn. Ponadto 40 osób to blondyni, a 60 to bruneci. Prawdopodobieństwo wylosowania osoby o jasnych włosach wynosi więc $P(Bl) = 0.4$.

a) Załóżmy, że wśród kobiet jest 25 blondynek i 30 brunetek. Jeśli losując 1 osobę dowiemy się najpierw, że to kobieta, zdarzenie K, wtedy szansa, że jest to blondynka wynosi $\frac{25}{55} \approx 0.45$, a więc wiedza o tym, że zaszło zdarzenie K zmienia niepewność co do zajścia zdarzenia Bl,

$$P(Bl|K) = \frac{P(Bl \cap K)}{P(K)} = \frac{25/100}{55/100} = 0.4545... \neq 0.4 = P(Bl)$$

Zdarzenia Bl i K są *zależne*.

b) Gdyby wśród kobiet były 22 blondynki i 33 brunetki, wtedy szansa na wylosowanie blondynki wśród pań wynosiłaby $\frac{22}{55} = 0.4$, a więc byłaby taka sama jak w ogólnej populacji. Wtedy wiedza o tym że wylosowana osoba jest kobietą nie zmienia niepewności co do tego, czy osoba ta ma jasne włosy. Zdarzenia Bl i K są tym razem *niezależne*.

$$P(Bl|K) = \frac{P(Bl \cap K)}{P(K)} = \frac{22/100}{55/100} = 0.4 = P(Bl) = \frac{40}{100}$$

.

Niezależność $P(A|B) = P(A)$ oznacza także, że zachodzi $P(A \cap B) = P(A) \cdot P(B)$, a więc także $P(B|A) = P(B)$.

**UWAGA:** Zdarzenia rozłączne nigdy nie są niezależne! Gdyby były niezależne, to co najmniej jedno z nich musiałoby być niemożliwe

$$P(A \cap B) = 0 \qquad \Leftrightarrow \qquad P(A) = 0 \ \vee \ P(B) = 0.$$

Gdyby wśród mężczyzn nie było ani jednego blondyna, $M \cap Bl = \emptyset$, wtedy $P(M|Bl) = 0 \neq P(M)$.

## Informacja

Powiedzmy, że w wyniku eksperymentu losowego zachodzi zdarzenie $A \in \Omega$. Informacja $I(A)$ mierzy jak bardzo to zajście zmniejsza naszą początkową niepewność co do wyniku doświadczenia. Jeśli zdarzenie $A$ jest pewne, t.j. zachodzi z prawdopodobieństwem 1, niepewność nie ulega zmianie, czyli

1. jeśli $P(A) = 1$, wtedy $I(A) = 0$.

Im mniej prawdopodobne było zdarzennie $A$, tym znaczniej ubywa niepewności na skutek jego zajścia, a więc

2. Jeśli $P(A) \leq P(B)$, wtedy $I(A) \geq I(B)$.

Jedoczesne zajście dwóch zdarzeń niezależnych zmniejsza niepewność o sumę informacji niesioną przez każde z tych zdarzeń

3. $I(A \cap B) = I(A) + I(B)$ jeśli $A$ i $B$ są niezależne, $P(A \cap B) = P(A)P(B)$.

Z tych postulatów wynika następująca postać informacji

$$I(A) = -\log P(A)\,.$$

Baza logarytmu jest dowolna (określa jednostkę informacji). Konwencjonalnie przyjmujemy bazę 2. Wówczas jednostkowa informacja nazywana jest bitem. Np. wynik rzutu symetryczną monetą niesie ze sobą 1 bit informacji,

$$I(O) = I(R) = -\log_2 0.5 = 1.$$

**Entropia Shannona dyskretnej zmiennej losowej**

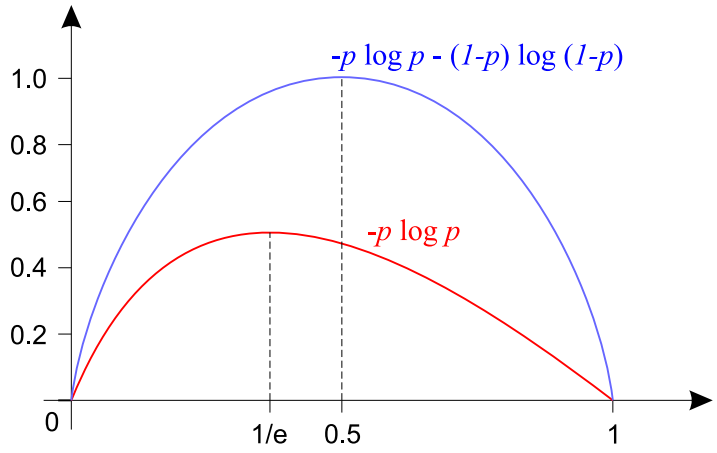
$$H(X) \;=\; \mathbb{E}(I(X)) \;=\; -\sum_{i=1}^{n} P(X=x_i)\log P(X=x_i) \;=\; -\sum_{i=1}^{n} p_i \log p_i.$$

Z racji tego, że $\lim\limits_{x\to 0} x\log x \;=\; 0$, przyjmujemy, że jeśli $p_i \;=\; 0$, odpowiadające temu zdarzenie elementarne nie wnosi nic do entropii rozkładu.

**Fakt** $H(X) \geq 0$. Ponadto $H(X) = 0$ wtedy i tylko wtedy gdy zmienna $X$ jest prawie wszędzie stała. $H$ przyjmuje maksimum równe $\log n$ dla zmiennej $X$ o rozkładzie równomiernym $p_i = \frac{1}{n}$, $i = 1, 2, ..., n$.

Przykład. $X \in \{0, 1\}$, $p_0 = p,\ p_1 = 1 - p$. Wówczas

$$H(X) = -p\log p - (1-p)\log(1-p) = h(p)$$



**Nierówność Gibbsa**

Jeśli $X$ jest zmienną losową przyjmującą wartości $x_1$, $x_2$, ..., $x_n$ z prawdopodobieństwami $p_1$, $p_2$, ..., $p_n$, $\sum p_i = 1$, oraz jeśli $q_1$, $q_2$, ..., $q_n \geq 0$ są innym układem liczb takich, że $\sum q_i = 1$, wówczas

$$-\sum p_i \log p_i \;\leq\; -\sum p_i \log q_i \,,$$

przy czym równość zachodzi wtedy i tylko wtedy gdy $p_i = q_i$, $i = 1, 2, ..., n$.

DOWÓD: Nierówność jest równoważna $\sum p_i \log \dfrac{q_i}{p_i} \;\leq\; 0$

Ponieważ $\log_2 c = \log_2 e \, \ln c$, możemy zamienić podstawę logarytmu na $e$

$$\log_2 e \;\sum p_i \ln \frac{q_i}{p_i} \;\leq\; 0$$

podzielić obie strony przez dodatni czynnik $\log_2 e$ i stosując nierówność $\ln t \leq t - 1$ otrzymamy

$$\sum p_i \ln \frac{q_i}{p_i} \leq \sum p_i \left( \frac{q_i}{p_i} - 1 \right) = \sum (q_i - p_i) = 0.$$

Ponieważ $\ln t = t - 1$ tylko gdy $t = 1$, nierówność powyżej przechodzi w równość jedynie gdy $\dfrac{q_i}{p_i} = 1$ dla wszystkich $i$.

**Entropia łączna**

Łączny rozkład wielu zmiennych losowych $\qquad P_{XY}(X = x_i, Y = y_j) = p_{ij}$

Zmienne $X, Y$ mogą być niezależne lub nie, $\qquad p_{ij} = p_i p_j \quad \Leftrightarrow \quad X \perp Y.$

Rozkłady brzegowe: $\qquad p_{i\cdot} = \sum_j p_{ij}, \quad p_{\cdot j} = \sum_i p_{ij}$

**Definicja** Entropia łączna

$$H(X, Y) = -\sum_{i,j} P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j)$$

**Twierdzenie** $H(X, Y) \leq H(X) + H(Y)$, równość jedynie gdy $X \perp Y$.

DOWÓD: Zauważmy, że $\sum_{i,j} p_{i\cdot}\, p_{\cdot j} = 1$.

$$\begin{aligned} H(X,Y) \;&=\; -\sum_{i,j} p_{ij} \log p_{ij} \;\leq\; -\sum_{i,j} p_{ij} \log(p_{i\cdot}\, p_{\cdot j}) \\ &=\; -\sum_{i,j} p_{ij} \Big( \log p_{i\cdot} + \log p_{\cdot j} \Big) \\ &=\; -\sum_i p_{i\cdot} \log p_{i\cdot} \;-\; \sum_j p_{\cdot j} \log p_{\cdot j} \;=\; H(X) + H(Y)\,, \end{aligned}$$

przy czym do równości potrzeba by zachodziło $p_{ij} = p_{i\cdot}\, p_{\cdot j}$ dla wszystkich $i, j$.

**Entropia warunkowa**

Prawdopodobieństwa warunkowe z rozkładu łącznego:

$$P(Y = y_j|X = x_i) \;=\; \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} \;=\; \frac{p_{ij}}{p_{i\cdot}}$$

$$\begin{aligned}
H(X,Y) &= -\sum_{i,j} P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) \\
H(X) &= -\sum_{i,j} P(X = x_i, Y = y_j) \log P(X = x_i) \\
H(X,Y) - H(X) &= -\sum_{i,j} P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} \\
&= -\sum_{i,j} P(X = x_i) P(Y = y_j|X = x_i) \log P(Y = y_j|X = x_i) \\
&= \sum_{i} P(X = x_i) H(Y|X = x_i) \\
&= H(Y|X)
\end{aligned}$$

Mamy więc $0 \leq H(Y|X) = H(X,Y) - H(X)$ z równością jedynie gdy $H(Y|X = x_i) = 0$ dla wszystkich $i$, to znaczy gdy $Y$ jest dokładnie określone na podstawie wartości $X$, czyli $Y = f(X)$. Mamy także

$$H(X) \leq H(X,Y) \leq H(X) + H(Y).$$

Przykład 1: Seria niezależnych rzutów symetryczną monetą, $X_i \in \{O, R\}$

$$P(X_1 = x_1, \ldots, X_n = x_n) \;=\; P(X_1 = x_1) \cdots P(X_n = x_n) \;=\; \left(\frac{1}{2}\right)^n$$

Więc $H(X_1, \ldots, X_n) = n \log_2 2 = n$. Rzuty kostką $H(...) = n \log_2 6 \approx n\, 2.59$.

Przykład 2: Na pasku w TV pojawiają się kolejne litery komunikatu, $X_i = i$-ta litera. Duża zależność (kolejne litery są łatwe do przewidzenia)

$$P(X_n = x_n|X_1 = x_1, \ldots X_{n-1} = x_{n-1}) \neq P(X_n = x_n)$$

Wtedy $H(X_1, \ldots, X_n) < nH(X)$.

**Kody bezprefiksowe**

$$c : \mathcal{A} \to \mathcal{B}^*, \qquad c^* : \mathcal{A}^* \to \mathcal{B}^* \qquad c^*(a_1 \dots a_k) = c(a_1) \dots c(a_k)$$

**Definicja** $c$ jest dekodowalny jeśli $c^*$ jest funkcją 1-1.

Różnowartościowość $c$ jest warunkiem koniecznym ale nie dostatecznym dekodowalności. Np.

$$c(0) = 0, \quad c(1) = 00, \quad c(2) = 000, \quad \dots \quad c(9) = 0000000000$$

ale $(c^*)^{-1}(00000000000000) = ?$

Słowo $\alpha$ jest prefiksem (przedrostkiem) słowa $\beta$,

$$\alpha \lhd \beta \quad \Leftrightarrow \quad \exists\, \gamma \quad \beta = \alpha\gamma$$

Np. $abba \lhd abbaaabbab$

**Definicja** Kod $c$ jest bezprefiksowy $\Leftrightarrow$ $\forall\, a, b \in \mathcal{A} \quad c(a) \not\lhd c(b)$

Przykład: $c(v) = 0, \quad c(x) = 10, \quad c(y) = 110, \quad c(z) = 111.$

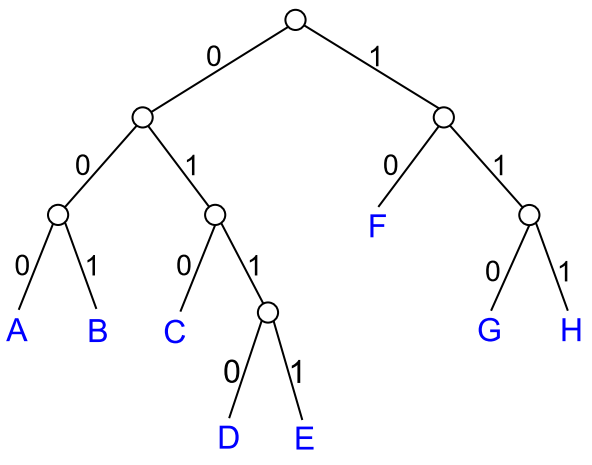
$$10001101111110100 1100 \quad \longrightarrow \quad 10\ 0\ 0\ 110\ 111\ 111\ 0\ 10\ 0\ 110\ 0$$

Kod dekodowalny nie musi być bezprefiksowy:

$$\begin{array}{lcl@{\qquad\qquad}lcl} c(x) & = & 10 & d(x) & = & 0 \\ c(y) & = & 100 & d(y) & = & 10 \\ c(z) & = & 1000 & d(z) & = & 100 \end{array}$$

Zdekodować słowo 100101000.

Kod bezprefiksowy pozwala na dekodowanie "w locie". Można go reprezentować za pomocą drzewa.



**Twierdzenie:** Nierówność Krafta I
Jeśli $c : \mathcal{A} \to \mathcal{B}^*$ jest kodem bezprefiksowym, dla którego słowa kodowe $c(a)$ mają długości $l(a)$ oraz $D$ jest liczbą liter w $\mathcal{B}$, wówczas

$$\sum_{a\in\mathcal{A}} D^{-l(a)} \;\leq\; 1\,.$$

**Twierdzenie:** Nierówność Krafta II
Niech $\mathcal{A}$ i $\mathcal{B}$ będą dwoma alfabetami, $|\mathcal{B}| = D$, oraz niech dla układu liczb $l(a)$, $a \in \mathcal{A}$, zachodzi nierówność

$$\sum_{a\in\mathcal{A}} D^{-l(a)} \;\leq\; 1\,.$$

Wówczas istnieje kod bezprefiksowy $c : \mathcal{A} \to \mathcal{B}^*$ o słowach kodowych $c(a)$ długości $l(a)$, $a \in \mathcal{A}$.

**Twierdzenie McMillana:** Jeśli $c : \mathcal{A} \to \mathcal{B}^*$ jest dekodowalny, $|\mathcal{B}| = D$, a słowa kodowe $c(a)$ mają długości $l(a)$, $a \in \mathcal{A}$, wówczas

$$\sum_{a\in\mathcal{A}} D^{-l(a)} \;\leq\; 1\,.$$

Dowód. Niech $L = \max_a l(a)$.

$$\left(\sum_{a\in\mathcal{A}} D^{-l(a)}\right)^R = \sum_{a_1,a_2,\ldots,a_R\in\mathcal{A}} D^{-l(a_1)-l(a_2)-\ldots-l(a_R)}$$

Oznaczmy

$$\omega = c^*(a_1a_2...a_R) = c(a_1)c(a_2)...c(a_R) \in \mathcal{B}^*\,, \qquad |\omega| = l(a_1)+l(a_2)+\cdots l(a_R) \leq LR\,.$$

Ponieważ kod $c$ jest z założenia dekodowalny, dowolnemu $\omega \in B^*$ odpowiada co najwyżej jedno słowo $a_1a_2...a_R$, takie, że $c^*(a_1a_2...a_R) = \omega$.

Przez $N(m)$ oznaczmy liczbę słów $\omega$ długości $m$, które są obrazami pewnych ciągów $a_1a_2...a_R$, $c^*(a_1a_2...a_R) = \omega$

$$\left(\sum_{a\in\mathcal{A}} D^{-l(a)}\right)^R \;\leq\; \sum_{m=1}^{LR} N(m)D^{-m} \;\leq\; \sum_{m=1}^{LR} D^m D^{-m} \;=\; LR\,,$$

a stąd

$$\sum_{a\in\mathcal{A}} D^{-l(a)} \;\leq\; \sqrt[R]{LR} \;\to\; 1 \quad \text{gdy} \quad R \to \infty\,.$$