13. cvičení z PSI

15. - 19. prosince 2014

13.1 (Hledání kódů) Pro informační zdroj zadaný pravděpodobnostmi

$$p_X(a) = 0.15,\ p_X(b) = 0.25,\ p_X(c) = 0.15,\ p_X(d) = 0.45$$

nalezněte binární Shannonův kód, binární Huffmanův kód a srovnejte jejich střední kódové délky.

Řešení:
Máme tedy náhodnou veličinu $X : \Omega \to \{a, b, c, d\}$. Pro konstrukci Shannonova binárního kódu $C_S : \{a, b, c, d\} \to \{0, 1\}^*$ potřebujeme znát délky jeho slov. Ty jsou dány vztahem $\ell(C_S(z)) = \lceil \log_2 \frac{1}{p_X(z)} \rceil$ pro $z \in \{a, b, c, d\}$.

Tedy $\ell(C_S(a)) = \ell(C_S(c)) = \lceil \log_2 100/15 \rceil = 3$, $\ell(C_S(b)) = \lceil \log_2 4 \rceil = 2$, $\ell(C_S(d)) = \lceil \log_2 100/45 \rceil = 2$. Po použití příslušného algoritmu (pro binární strom hloubky $\max\{\ell(C_S(z)) \mid z \in \{a, b, c, d\}\} = 3$) dostaneme např. kód

$$C_S(a) = 000,\ C_S(b) = 01,\ C_S(c) = 001,\ C_S(d) = 10.$$

Jeho střední délka pak je $L(C_S) = E(\ell(C_S(X))) = \sum_{z \in \{a,b,c,d\}} p_X(z) \ell(C_S(z)) = 2 \cdot 0,15 \cdot 3 + 0,25 \cdot 2 + 0,45 \cdot 2 = 2,3$.

Huffmanův binární kód $C_H : \{a, b, c, d\} \to \{0, 1\}^*$ pak pomocí příslušného algoritmu může vypadat takto:

$$C_H(a) = 000,\ C_H(b) = 01,\ C_H(c) = 001,\ C_H(d) = 1.$$

Jeho střední délka pak je $L(C_H) = E(\ell(C_H(X))) = \sum_{z \in \{a,b,c,d\}} p_X(z) \ell(C_H(z)) = 2 \cdot 0,15 \cdot 3 + 0,25 \cdot 2 + 0,45 \cdot 1 = 1,85$. Tedy $L(C_S) = 2,3 > 1,85 = L(C_H)$.

13.2 (Přenos bitů) Uvažujme binární kanál s pravděpodobností chyby $p = 10^{-4}$ při přenosu jakéhokoliv bitu. Jaká je pravděpodobnost chyby v dekódování, pokud vstupní bit pošleme jako zprávu, kde daný bit zopakujeme $3\times$ a dekódujeme ho pak jako znak, který má ve výsledné přijaté zprávě větší zastoupení? Jaký je střední počet chyb ve výsledné zprávě?

Řešení:
Předpokladáme, že jednotlivé chyby ve zprávě jsou nezávislé, tj. veličiny X_i vyjadřující počet chyb na i-tém místě zprávy jsou nezávislé pro $i = 1, 2, 3$ a mají alternativní rozdělení. Tedy $P(X_i = 1) = p$ a $P(X_i = 0) = 1 - p$. Celkový počet chyb ve zprávě

je tedy veličina $Y = \sum_{i=1}^{3} X_i$ a má binomické rozdělení $Bi(3, p)$. Pravděpodobnost chyby při dekódování tedy je $P(Y \geq 2) = p^3 + 3p^2(1-p) \doteq 3 \cdot 10^{-8}$. Střední hodnota počtu chyb ve zprávě pak je $E(Y) = \sum_{i=1}^{3} E(X_i) = 3p = 3 \cdot 10^{-4}$.

13.3 (Rychlost entropie) Žába skáče mezi třemi kameny. V každém kroku skočí na jeden ze dvou sousedních kamenů s pravděpodobností 1/2. Určete rychlost entropie stacionárního procesu, který zaznamenává její pohyb.

Řešení:
Jedná se o markovský řetězec s maticí

$$\mathbf{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Stacionární rozdělení je $\vec{p} = (1/3, 1/3, 1/3)$. Rychlost entropie se pak počítá podle vzorce:

$$H\big((X_n)_{n \in \mathbb{N}}\big) = H(X_2|X_1) = \sum_{i=1}^{3} p_i H(X_2|X_1 = i),$$

kde p_i je i-tá složka vektoru $\vec{p}$. Dále je

$$H(X_2|X_1 = i) = -\sum_{j=1}^{3} P(X_2 = j|X_1 = i) \cdot \log P(X_2 = j|X_1 = i) =$$

$$-\sum_{j=1}^{3} p_{i,j} \cdot \log p_{i,j} = 0 \log 0 + \frac{1}{2} \log 2 + \frac{1}{2} \log 2 = 1$$

kde $p_{i,j}$ je příslušný prvek matice $\mathbf{P}$.

Celkem tedy dostáváme, že $H\big((X_n)_{n \in \mathbb{N}}\big) = \sum_{i=1}^{3} p_i \cdot 1 = 1$.