SHANNONOVA TEORIE TAJNÉ KOMUNIKACE

Verze 1.4

ANDREW KOZLÍK

1. ÚVOD DO DISKRÉTNÍ PRAVDĚPODOBNOSTI

Uvažujme náhodný pokus, při kterém házíme dokonale symetrickou kostkou a zaznamenáváme číslo, které padlo. S každým náhodným pokusem je spojena množina všech možných výsledků pokusu. Tato množina se nazývá *množina elementárních jevů* a značí se Ω. V našem případě $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Definice. Nechť Ω je spočetná množina a $\Pr : \Omega \to [0, 1]$ je zobrazení takové, že $\sum_{\omega\in\Omega} \Pr(\omega) = 1$. Potom říkáme, že $(\Omega, \Pr)$ je *diskrétní pravděpodobnostní prostor*, prvky množiny Ω nazýváme *elementární jevy* a $\Pr(\omega)$ nazýváme *pravděpodobnost* elementárního jevu ω.

V případě náhodného hodu kostkou máme $\Pr(\omega) = \frac{1}{6}$ pro všechna $\omega \in \Omega$.

Definice. Nechť $(\Omega, \Pr)$ je diskrétní pravděpodobnostní prostor. Libovolná podmnožina $E \subseteq \Omega$ se nazývá *(náhodný) jev* a definujeme $\Pr(E) = \sum_{\omega\in E} \Pr(\omega)$ a $\Pr(\emptyset) = 0$. Oddělujeme-li více jevů čárkami, rozumíme tím jejich průnik $\Pr(E_1, E_2, \ldots, E_n) = \Pr(E_1 \cap \cdots \cap E_n)$.

Chceme-li znát pravděpodobnost, že na kostce padne sudé číslo, pak tento jev vyjádříme jako $E = \{2, 4, 6\}$ a spočteme $\Pr(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. Tuto úvahu můžeme vyjádřit také pomocí tzv. náhodné veličiny.

Definice. Nechť $(\Omega, \Pr)$ je diskrétní pravděpodobnostní prostor, A je konečná množina a $X : \Omega \to A$ je zobrazení. Potom říkáme, že X je *diskrétní náhodná veličina*, a pro každé $a \in A$ definujeme jev $\{\, \omega \in \Omega : X(\omega) = a \,\}$, který značíme $X{=}a$. Máme tedy

$$\Pr(X{=}a) = \Pr(\{\, \omega \in \Omega : X(\omega) = a \,\}) = \sum_{\omega : X(\omega)=a} \Pr(\omega).$$

Definujme zobrazení $X : \Omega \to \{0, 1\}$, které každému možnému výsledku hodu kostkou ω přiřazuje $\omega \bmod 2$. Potom $\Pr(X{=}0) = \Pr(\{2, 4, 6\}) = \frac{1}{2}$ udává pravděpodobnost, že na kostce padne číslo, jehož zbytek po dělení 2 bude 0, jinými slovy sudé číslo.

Definice. Nechť $(\Omega, \Pr)$ je diskrétní pravděpodobnostní prostor a E_1, $E_2 \subseteq \Omega$ jsou náhodné jevy takové, že $\Pr(E_2) \neq 0$. Potom definujeme *podmíněnou pravděpodobnost*

$$\Pr(E_1 \mid E_2) := \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)}$$

(čteme „pravděpodobnost E_1 za podmínky E_2“).

Uvažujme náhodný pokus, při kterém dvakrát po sobě házíme kostkou. Máme tedy $\Omega = \{1,\dots,6\}^2$. Spočítáme pravděpodobnost, že rozdíl hodnot na kostkách je 2, za podmínky, že součet hodnot na kostkách je roven 8. Máme

$$E_1 = \{(1,3),(3,1),(2,4),(4,2),(3,5),(5,3),(4,6),(6,4)\},$$
$$E_2 = \{(2,6),(6,2),(3,5),(5,3),(4,4)\} \quad \text{a} \quad E_1 \cap E_2 = \{(3,5),(5,3)\},$$
$$\text{čili} \quad \Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)} = \frac{2 \cdot \frac{1}{36}}{5 \cdot \frac{1}{36}} = \frac{2}{5}.$$

Definice. Říkáme, že diskrétní náhodná veličina $X : \Omega \to A$ má *rovnoměrné rozdělení*, jestliže $\Pr(X{=}a) = |A|^{-1}$ pro všechna $a \in A$.

Definice. Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Říkáme, že X a Y jsou *nezávislé*, jestliže

$$\Pr(X{=}a, Y{=}b) = \Pr(X{=}a)\Pr(Y{=}b) \quad \text{pro všechna } a \in A \text{ a } b \in B.$$

Věta 1.1 (o úplné pravděpodobnosti)**.** *Nechť $E_1,\dots,E_n \subseteq \Omega$ jsou po dvou disjunktní jevy takové, že $\bigcup_{i=1}^n E_i = \Omega$. Potom pro libovolný jev $E \subseteq \Omega$ platí*

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E, E_i).$$

Jsou-li navíc $\Pr(E_i) > 0$ pro všechna $i \in \{1,\dots,n\}$, pak můžeme psát

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E \mid E_i)\Pr(E_i).$$

Důkaz. Sdruženou pravděpodobnost $\Pr(E, E_i)$ rozepíšeme podle definice a využijeme toho, že $(E \cap E_i)$ a $(E \cap E_j)$ jsou disjunktní pro $i \neq j$ a že $\bigcup_{i=1}^n (E \cap E_i) = E \cap \Omega = E$.

$$\sum_{i=1}^{n} \Pr(E, E_i) = \sum_{i=1}^{n} \sum_{\omega \in E \cap E_i} \Pr(\omega) = \sum_{\omega \in E} \Pr(\omega) = \Pr(E).$$

Druhý vzorec je jen přepisem prvního vzorce podle definice podmíněné pravděpodobnosti. □

Důsledek 1.2. *Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny a $a \in A$. Potom*

$$\Pr(X{=}a) = \sum_{y \in B} \Pr(X{=}a, Y{=}b).$$

Definice. Nechť $X : \Omega \to A$ je diskrétní náhodná veličina, kde $A \subset \mathbb{R}$. Potom definujeme *střední hodnotu*

$$\mathrm{E}\, X = \sum_{\omega \in \Omega} \Pr(\omega) X(\omega).$$

Máme-li náhodnou veličinu X, která udává výsledek hodu kostkou, pak její střední hodnota je $\mathrm{E}\, X = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = 3{,}5$.

Tvrzení 1.3. *Nechť $X : \Omega \to A$ je diskrétní náhodná veličina a $f : A \to \mathbb{R}$. Potom*

$$\mathrm{E}\, f(X) = \sum_{a \in A} \Pr(X{=}a) f(a).$$

Důkaz.

$$\sum_{a\in A} \Pr(X{=}a) f(a) = \sum_{a\in A} \sum_{\omega: X(\omega)=a} \Pr(\omega) f(a) = \sum_{\omega\in\Omega} \Pr(\omega) f(X(\omega)) = \mathrm{E}\, f(X).$$

□

Příklad 1.4. Nechť $X : \Omega \to A$ je diskrétní náhodná veličina s rovnoměrným rozdělením na $A = \{-2, -1, 0, 1, 2\}$. Nemusíme znát Ω, abychom spočítali $\mathrm{E}\, X$. Stačí v předchozím tvrzení položit $f = \mathrm{id}$ a máme $\mathrm{E}\, X = \frac{1}{5}\cdot(-2)+\frac{1}{5}\cdot(-1)+\frac{1}{5}\cdot 0+\frac{1}{5}\cdot 1+\frac{1}{5}\cdot 2$. Zajímá-li nás náhodná veličina X^2, pak její střední hodnotu můžeme spočítat například jedním z následujících dvou způsobů. Můžeme vycházet z jejího rozdělení pravděpodobnosti $\Pr(X^2{=}0) = \frac{1}{5}$, $\Pr(X^2{=}1) = \frac{2}{5}$ a $\Pr(X^2{=}4) = \frac{2}{5}$, čímž dostaneme $\mathrm{E}\, X^2 = \frac{1}{5}\cdot 0+\frac{2}{5}\cdot 1+\frac{2}{5}\cdot 4 = 2$. Jednodušší je však přímočaře dosadit v předchozím tvrzení za f zobrazení $x \mapsto x^2$, čili $\mathrm{E}\, X^2 = \frac{1}{5}\cdot(-2)^2 + \frac{1}{5}\cdot(-1)^2 + \frac{1}{5}\cdot 0^2 + \frac{1}{5}\cdot 1^2 + \frac{1}{5}\cdot 2^2 = 2$.

Cvičení 1.5. (a) Ověřte, že pro libovolné jevy A, $B \subseteq \Omega$ platí $\Pr(A\cup B) = \Pr(A) + \Pr(B) - \Pr(A, B)$.

(b) Ověřte, že pro libovolný jev $A \subseteq \Omega$ platí, že pravděpodobnost, že jev A nenastane, je $\Pr(\Omega \setminus A) = 1 - \Pr(A)$.

(c) Z balíčku 108 kanastových karet vybereme náhodně jednu kartu. Jaká je pravděpodobnost, že jsme vybrali srdcovou kartu nebo sedmičku?

(d) Jaká je pravděpodobnost, že jsme nevybrali ani srdcovou kartu, ani sedmičku?

2. ENTROPIE A HUFFMANOVO KÓDOVÁNÍ

Uvažujme experiment, při kterém k-krát po sobě hodíme mincí a výsledky hodů zaznamenáme jako posloupnost hodnot 0 (rub) a 1 (líc). Dostáváme tak posloupnost z množiny $\{0,1\}^k$, přičemž každá posloupnost v této množině má pravděpodobnost 2^{-k}. Tento zápis je zřejmě nejkompaktnější mezi všemi zápisy, které využívají symboly 0 a 1. Můžeme říct, že informační hodnota takového zápisu je k bitů. Toto lze formulovat také tak, že náhodná veličina X s rovnoměrným rozdělením na množině $\{0,1\}^k$ má míru nejistoty k bitů.

Nyní pro změnu uvažujme experiment, při kterém k-krát po sobě házíme dvěma nerozlišitelnými mincemi současně. Opět zaznamenáme výsledky hodů, a jelikož mince jsou nerozlišitelné, sledujeme následující výsledky: (rub a rub), (rub a líc) nebo (líc a líc), s pravděpodobnostmi $\frac{1}{4}$, $\frac{1}{2}$ a $\frac{1}{4}$. Nyní bychom chtěli zaznamenat výsledky hodů a chtěli bychom k tomu využít stejný aparát jako v předchozím případě, tj. posloupnost nul a jedniček. Dva symboly nám pochopitelně nestačí k zaznamenání tří možných výsledků, ale každý hod můžeme zakódovat dvojicí symbolů 00, 01 nebo 11 a tyto napsat do posloupnosti po sobě. Výsledný záznam tak bude mít délku $2k$. Existuje však efektivnější způsob. Jednak si všimněme, že nevyužíváme plně kapacitu daného aparátu, protože kód 10 se nepoužije. To znamená, že například posloupnost 01**10**01 je neplatná. Dále si všimněme, že výsledek (rub a líc) se bude vyskytovat dvakrát častěji než ostatní výsledky a tento fakt by bylo dobré v zápisu zohlednit tím, že bude mít kratší zápis. Alternativní zápis by mohl vypadat tak, že pro výsledek (rub a rub) zvolíme kód 00, pro (rub a líc) 1 a pro (líc a líc) 01. Tyto výsledky pak budeme psát do posloupnosti po sobě tak, jak budou padat. Všimněme si, že i když výsledky budou reprezentovány kódy různých délek, půjde poznat, kde každý kód končí. Víme totiž, že kód začínající symbolem 0 má vždy délku dvou symbolů a kód začínající symbolem 1 má délku jednoho symbolu. Například 011001 má nyní význam (líc a líc), (rub a líc), (rub a rub) a (rub a líc). Z každé posloupnosti lze tedy jednoznačně rekonstruovat výsledky hodů. Říkáme, že toto kódování je *prosté*. Délka výsledné posloupnosti nebude jednoznačně určena počtem hodů k, ale víme, že bude ležet v intervalu $[k, 2k]$ a průměrná délka záznamu bude

$\frac{1}{4}k \cdot 2 + \frac{1}{2}k \cdot 1 + \frac{1}{4}k \cdot 2 = \frac{3}{2}k$. Dá se ukázat, že toto kódování je z hlediska délky optimální. Průměrné množství informace získané z jednoho hodu dvěma nerozlišitelnými mincemi je tedy 1,5 bitů.

Ve svém průlomovém článku [1] z roku 1948 zavedl Claude Shannon pojem informační entropie diskrétní náhodné veličiny. Mimo jiné dokázal, že entropie náhodné veličiny je dolní mez na průměrnou délku jejího libovolného prostého kódování. Zároveň dokázal, že vždy existuje prosté kódování, kterým se lze přiblížit k této dolní mezi libovolně blízko. Entropie tedy odpovídá našemu chápání míry nejistoty o výsledku experimentu neboli průměrnému množství informace obsaženému ve výsledku experimentu.

Definice. Nechť $X : \Omega \to A$ je diskrétní náhodná veličina na diskrétním pravděpodobnostním prostoru $(\Omega, \Pr)$. Potom *entropie* náhodné veličiny X je definována jako

$$\mathrm{H}(X) = -\sum_{a \in A} \Pr(X{=}a) \cdot \log_2 \Pr(X{=}a),$$

kde předepisujeme $0 \cdot \log_2 0 := 0$.

Předpis $0 \cdot \log_2 0 = 0$ v předchozí definici je ospravedlněn jednak limitním chováním příslušné funkce $\lim_{z \to 0_+} z \log_2 z = 0$, jednak tím, že je-li $X{=}a$ nemožný jev, pak takový jev by nijak neměl zvyšovat míru nejistoty náhodné veličiny. Pro lepší čitelnost budeme při počítání s entropií mlčky předpokládat, že každá náhodná veličina $X : \Omega \to A$ splňuje $\Pr(X{=}a) \neq 0$ pro všechna $a \in A$. V opačném případě by totiž jednoduše stačilo definovat novou náhodnou veličinu X', která tuto vlastnost splňuje a přitom $\mathrm{H}(X) = \mathrm{H}(X')$.

Všimněme si, že má-li náhodná veličina X rovnoměrné rozdělení na n-prvkové množině, pak $\mathrm{H}(X) = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n$. V případě $n = 2^k$ tak dostáváme $\mathrm{H}(X) = k$.

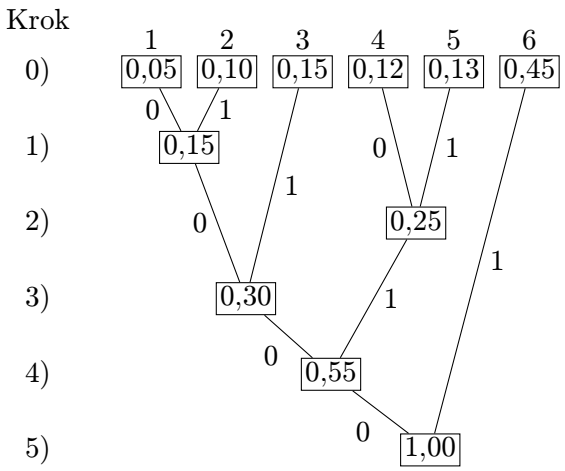
Dále si všimněme, že existuje-li $a_0 \in A$ takové, že $\Pr(X{=}a_0) = 1$, pak $H(X) = 1 \cdot \log_2 1 = 0$. Jinými slovy, je-li výsledek experimentu předem jistý, pak entropie je nulová.

Abychom nahlédli, odkud se vzorec pro entropii bere, ukážeme si, jak sestrojit tzv. Huffmanovo kódování náhodné veličiny $X : \Omega \to A$. Toto kódování je prosté a dá se ukázat, že průměrná délka zakódovaného prvku z A leží v intervalu $[\mathrm{H}(X), \mathrm{H}(X) + 1)$. S Huffmanovým kódováním jsme se již setkali v posledním příkladu o házení mincemi, ve kterém se dokonce podařilo dosáhnout dolní meze $\mathrm{H}(X) = 1{,}5$.

Sestrojení Huffmanova kódování nejlépe uvidíme na příkladu. Mějme náhodnou veličinu $X : \Omega \to A$, kde $A = \{1, 2, 3, 4, 5, 6\}$, s pravděpodobnostním rozdělením

a	1	2	3	4	5	6
$\Pr(X{=}a)$	0,05	0,10	0,15	0,12	0,13	0,45

Sestrojíme binární strom, jehož listy budou prvky A. Algoritmus funguje tak, že postupně spojuje menší stromy do větších, až na konci získáme jediný strom, z něhož vyčteme Huffmanovy kódy listů. Pro každý list $a \in A$ definujeme jeho váhu jako $\Pr(X{=}a)$. Na začátku algoritmu máme množinu vrcholů A a na každý z nich můžeme nahlížet jako na kořen jednoprvkového stromu. V každém kroku algoritmu najdeme dva nejlehčí kořenové vrcholy a vytvoříme nový kořenový vrchol, ke kterému je připojíme. Váhu nového vrcholu definujeme jako součet vah připojených vrcholů. Tímto krokem se počet kořenových vrcholů zmenší o jeden. Na konci algoritmu zbývá jediný kořenový vrchol, a tedy jediný strom. Z každého vrcholu vycházejí dvě větve směrem k jeho potomkům. Na závěr algoritmu projdeme vrcholy a jednu větev vždy označíme symbolem 0 a druhou symbolem 1. Toto označení můžeme provést náhodně. Na obrázku 2.1 máme strom příslušný náhodné veličině X se shora uvedeným rozdělením pravděpodobnosti. Z obrázku vidíme, ve kterém kroku algoritmu jednotlivé vrcholy vznikly. Huffmanův kód $h(a)$ libovolného prvku $a \in A$ zjistíme tak, že najdeme cestu od kořene stromu k příslušnému listu a přečteme symboly ležící na této cestě. Tímto získáme následující kódy:



Obrázek 2.1: Huffmanův algoritmus.

a	$h(a)$	$-\log_2 \Pr(X{=}a)$
1	0000	4,32
2	0001	3,32
3	001	2,74
4	010	3,06
5	011	2,94
6	1	1,15

Všimněme si, že pro každé $a \in A$ je délka kódu $\ell(h(a))$ přibližně rovna $-\log_2 \Pr(X{=}a)$. Střední hodnotu délky kódu spočítáme jako

$$\mathrm{E}\,\ell(h(X)) = \sum_{a \in A} \Pr(X{=}a) \cdot \ell(h(a)).$$

Vzhledem k tomu, že hodnoty $\ell(h(a))$ se přibližně rovnají hodnotám $-\log_2 \Pr(X{=}a)$, není nic překvapivého na tom, že i $\mathrm{E}\,\ell(h(X)) \approx \mathrm{H}(X)$.

$$\begin{aligned}\mathrm{E}\,\ell(h(X)) &= 0{,}05 \cdot 4 + 0{,}10 \cdot 4 + 0{,}15 \cdot 3 + 0{,}12 \cdot 3 + 0{,}13 \cdot 3 + 0{,}45 \cdot 1 = 2{,}25 \\ \mathrm{H}(X) &\approx 0{,}05 \cdot 4{,}32 + 0{,}10 \cdot 3{,}32 + 0{,}15 \cdot 2{,}74 + 0{,}12 \cdot 3{,}06 + 0{,}13 \cdot 2{,}94 + 0{,}45 \cdot 1{,}15 \approx 2{,}23\end{aligned}$$

3. ZÁKLADNÍ VLASTNOSTI ENTROPIE

Věta 3.1 (Jensenova nerovnost)**.** *Nechť $f : I \to \mathbb{R}$ je ryze konkávní funkce na intervalu I, $x_1, \ldots, x_n \in I$, a nechť $\lambda_1, \ldots, \lambda_n$ jsou kladná čísla taková, že $\sum_{i=1}^n \lambda_i = 1$. Potom*

$$\sum_{i=1}^{n} \lambda_i f(x_i) \le f\Big(\sum_{i=1}^{n} \lambda_i x_i\Big),$$

přičemž rovnost nastává právě tehdy, když $x_1 = x_2 = \cdots = x_n$.

Věta 3.2 (o maximální entropii)**.** *Nechť $X : \Omega \to A$ je diskrétní náhodná veličina, pak*

$$\mathrm{H}(X) \le \log_2 |A|,$$

přičemž rovnost nastává právě tehdy, když X má rovnoměrné rozdělení.

Důkaz. Využijeme Jensenovu nerovnost s $f = \log_2$

$$\mathrm{H}(X) = -\sum_{a\in A} \Pr(X{=}a)\cdot \log_2 \Pr(X{=}a) = \sum_{a\in A} \Pr(X{=}a)\cdot \log_2 \frac{1}{\Pr(X{=}a)} \le \log_2 \sum_{a\in A} 1.$$

□

Definice. Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Definujeme *kartézský součin* veličin X a Y jako náhodnou veličinu $(X, Y) : \Omega \to A \times B$ danou předpisem $(X, Y) : \omega \mapsto (X(\omega), Y(\omega))$. *Sdruženou entropii* veličin X a Y definujeme jako entropii jejich kartézského součinu a značíme ji $\mathrm{H}(X, Y)$.

Poznámka 3.3. Podle předchozí definice pro všechna $a \in A$ a $b \in B$ platí

$$\begin{aligned}\Pr((X,Y){=}(a,b)) &= \Pr(\{\,\omega\in\Omega : X(\omega)=a, Y(\omega)=b\,\}) \\ &= \Pr(\{\,\omega\in\Omega : X(\omega)=a\,\}\cap\{\,\omega\in\Omega : Y(\omega)=b\,\}) = \Pr(X{=}a, Y{=}b).\end{aligned}$$

Sdruženou entropii lze tedy rozepsat

$$\mathrm{H}(X,Y) = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a, Y{=}b)\log_2 \Pr(X{=}a, Y{=}b). \tag{3.1}$$

Všimněme si, že kartézský součin náhodných veličin a jejich sdruženou entropii lze přirozeným způsobem rozšířit na více veličin. V zápisu sdružené entropie pak nezáleží na pořadí veličin, ani na jejich případném uzávorkování, tj. platí například

$$\mathrm{H}(X,Y,Z) = \mathrm{H}(X,(Y,Z)) = \mathrm{H}((Z,X),Y).$$

Věta 3.4 (o sdružené entropii)**.** *Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Potom $\mathrm{H}(X, Y) \le \mathrm{H}(X) + \mathrm{H}(Y)$, přičemž rovnost nastává právě tehdy, když X a Y jsou nezávislé.*

Důkaz. Jelikož podle důsledku 1.2 je $\Pr(X{=}a) = \sum_{b\in B} \Pr(X{=}a, Y{=}b)$, můžeme rozepsat

$$\mathrm{H}(X) = -\sum_{a\in A} \Pr(X{=}a)\log_2 \Pr(X{=}a) = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a, Y{=}b)\log_2 \Pr(X{=}a).$$

Podobně také

$$\mathrm{H}(Y) = -\sum_{\substack{a\in A\\ b\in B}} \Pr(X{=}a, Y{=}b)\log_2 \Pr(Y{=}b). \tag{3.2}$$

Sčítance, ve kterých $\Pr(X{=}a, Y{=}b) = 0$, můžeme v těchto součtech i v rovnosti (3.1) vynechávat, což níže také činíme, abychom se vyhnuli dělení nulou. Důkaz provedeme tak, že ukážeme, že výraz $\mathrm{H}(X, Y) - \mathrm{H}(X) - \mathrm{H}(Y)$ není kladný a že je roven nule právě tehdy, když X a Y jsou nezávislé. Spojením shora uvedených rovností s rovnicí (3.1) dostáváme

$$\mathrm{H}(X,Y) - \mathrm{H}(X) - \mathrm{H}(Y) = \sum_{\substack{a\in A,\ b\in B\\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b)\log_2 \frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a, Y{=}b)}. \tag{3.3}$$

Aplikujeme Jensenovu nerovnost s $f = \log_2$

$$\begin{aligned}\mathrm{H}(X,Y) - \mathrm{H}(X) - \mathrm{H}(Y) &\le \log_2 \sum_{\substack{a\in A,\ b\in B\\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a)\Pr(Y{=}b) \le \log_2 \sum_{\substack{a\in A\\ b\in B}} \Pr(X{=}a)\Pr(Y{=}b) \\ &= \log_2 \sum_{a\in A}\left(\Pr(X{=}a)\sum_{b\in B}\Pr(Y{=}b)\right) = \log_2 \sum_{a\in A}\Pr(X{=}a) = \log_2 1 = 0.\end{aligned}$$

Jsou-li X a Y nezávislé, tj. $\Pr(X{=}a, Y{=}b) = \Pr(X{=}a)\Pr(Y{=}b)$, pak pravá strana rovnosti (3.3) je zřejmě nulová. Naopak, jsou-li strany rovnosti (3.3) nulové, pak v Jensenově nerovnosti nastává rovnost a podle věty 3.1 existuje $c \in \mathbb{R}$ takové, že $\frac{\Pr(X=a)\Pr(Y=b)}{\Pr(X=a,Y=b)} = c$ pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) \neq 0$. Aby v součtu vycházela 0, je jedinou možností $c = 1$, čili $\Pr(X{=}a, Y{=}b) = \Pr(X{=}a)\Pr(Y{=}b)$ pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) \neq 0$. Zbývá dokázat, že tato rovnost platí i pro zbývající a a b, tj. pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) = 0$, platí $\Pr(X{=}a)\Pr(Y{=}b) = 0$. Máme

$$\sum_{\substack{a\in A\\ b\in B}} \Pr(X{=}a)\Pr(Y{=}b) \;=\; 1 \;=\; \sum_{\substack{a\in A,\ b\in B\\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) \;=\; \sum_{\substack{a\in A,\ b\in B\\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a)\Pr(Y{=}b)$$

a odtud vidíme, že

$$\sum_{\substack{a\in A,\ b\in B\\ \Pr(X=a,Y=b)=0}} \Pr(X{=}a)\Pr(Y{=}b) = 0.$$

□

Definice. Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny a $E \subseteq \Omega$ je jev takový, že $\Pr(E) \neq 0$. Definujeme *entropii veličiny X podmíněnou jevem E*

$$\mathrm{H}(X \mid E) = -\sum_{a\in A} \Pr(X{=}a \mid E)\log_2 \Pr(X{=}a \mid E).$$

Podmíněnou entropii $\mathrm{H}(X \mid Y)$ definujeme jako vážený průměr $\mathrm{H}(X \mid Y{=}b)$ přes všechny možné hodnoty $b \in B$

$$\mathrm{H}(X \mid Y) = \sum_{b\in B} \Pr(Y{=}b)\,\mathrm{H}(X \mid Y{=}b).$$

Tvrzení 3.5. *Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny, potom*

$$\mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y).$$

Důkaz. Definici podmíněné entropie můžeme rozepsat

$$\begin{aligned}\mathrm{H}(X \mid Y) &= -\sum_{b\in B} \Pr(Y{=}b) \sum_{a\in A} \Pr(X{=}a \mid Y{=}b)\log_2 \Pr(X{=}a \mid Y{=}b)\\ &= -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a, Y{=}b)\log_2 \Pr(X{=}a \mid Y{=}b).\end{aligned}$$

Rozepíšeme-li entropii $\mathrm{H}(Y)$ stejně jako v rovnosti (3.2) z důkazu věty o sdružené entropii, pak spojením s předchozí rovností dostáváme

$$\mathrm{H}(X \mid Y) + \mathrm{H}(Y) = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a, Y{=}b)\log_2 \Pr(X{=}a \mid Y{=}b)\Pr(Y{=}b) = \mathrm{H}(X, Y).$$

□

Podle tvrzení 3.5 a věty 3.4 platí

$$0 \leq \mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y) \leq \mathrm{H}(X),$$

přičemž rovnost mezi $\mathrm{H}(X, Y) - \mathrm{H}(Y)$ a $\mathrm{H}(X)$ nastává právě tehdy, když X a Y jsou nezávislé náhodné veličiny. Tímto dostáváme dva důsledky.

Důsledek 3.6. *Pro libovolné diskrétní náhodné veličiny X a Y platí $\mathrm{H}(X \mid Y) \leq \mathrm{H}(X)$, přičemž rovnost nastává právě tehdy, když X a Y jsou nezávislé.*

Důsledek 3.7. *Pro libovolné diskrétní náhodné veličiny X a Y platí* $\mathrm{H}(Y) \leq \mathrm{H}(X,Y)$.

Definice. Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Definujeme *vzájemnou informaci*

$$I(X;Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y).$$

Všimněme si několika základních faktů o vzájemné informaci, které plynou z našich poznatků o entropii.

Tvrzení 3.8. *Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Potom*

1. $I(X;Y) = I(Y;X)$;
2. $I(X;Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y) = \mathrm{H}(Y) - \mathrm{H}(Y \mid X)$;
3. $I(X;Y) \geq 0$, *přičemž rovnost nastává právě tehdy, když X a Y jsou nezávislé.*

Příklad 3.9. Uvažujme experiment, při kterém házíme sedmkrát po sobě mincí a zaznamenáváme výsledky hodů. Definujme náhodnou veličinu X, která udává výsledky prvních 6 hodů, a náhodnou veličinu Y, která udává výsledky posledních 3 hodů. Zřejmě $\mathrm{H}(X) = 6$ a $\mathrm{H}(Y) = 3$. Chceme-li znát výsledky prvních 6 i posledních 3 hodů, pak potřebujeme úplnou informaci o výsledku experimentu, čili $\mathrm{H}(X,Y) = 7$. Nyní si položme otázku, známe-li pouze výsledek posledních 3 hodů, pak jakou mírou nejistoty je zatížen výsledek prvních 6 hodů? Schází nám znalost výsledků prvních 4 hodů, míra nejistoty je tedy 4 bity, což zapisujeme $\mathrm{H}(X \mid Y) = 4$. Známe-li naopak výsledek prvních 6 hodů, pak ke znalosti výsledku posledních 3 hodů nám schází pouze znalost posledního hodu, čili $\mathrm{H}(Y \mid X) = 1$. Vidíme, že vzorec z tvrzení 3.5 potvrzuje naše chápání podmíněné entropie. Dále vidíme, že náhodné veličiny mezi sebou sdílejí informaci o výsledku 5. a 6. hodu, což lze vyjádřit tak, že vzájemná informace těchto dvou veličin je $I(X;Y) = 2$.

4. VZDÁLENOST JEDNOZNAČNOSTI

Tvrzení 4.1. *Mějme šifru a nechť $\mathbf{K}$, $\mathbf{P}$ a $\mathbf{C}$ jsou náhodné veličiny, které odpovídají volbě klíče, volbě otevřeného textu a výslednému šifrovému textu. Potom*

$$\mathrm{H}(\mathbf{K} \mid \mathbf{C}) = \mathrm{H}(\mathbf{K}) + \mathrm{H}(\mathbf{P}) - \mathrm{H}(\mathbf{C}).$$

Důkaz. Jelikož znalost klíče a otevřeného textu jednoznačně určuje šifrový text, máme $\mathrm{H}(\mathbf{C} \mid (\mathbf{K},\mathbf{P})) = 0$. Stejně tak $\mathrm{H}(\mathbf{P} \mid (\mathbf{K},\mathbf{C})) = 0$. Veličiny $\mathbf{K}$ a $\mathbf{P}$ jsou nezávislé, proto $\mathrm{H}(\mathbf{K},\mathbf{P}) = \mathrm{H}(\mathbf{K}) + \mathrm{H}(\mathbf{P})$. Důkaz provedeme tím, že entropii $\mathrm{H}(\mathbf{K},\mathbf{P},\mathbf{C})$ rozepíšeme dvěma způsoby

$$\begin{aligned}\mathrm{H}(\mathbf{K},\mathbf{P},\mathbf{C}) &= \mathrm{H}(\mathbf{C} \mid (\mathbf{K},\mathbf{P})) + \mathrm{H}(\mathbf{K},\mathbf{P}) = \mathrm{H}(\mathbf{K}) + \mathrm{H}(\mathbf{P}),\\ \mathrm{H}(\mathbf{K},\mathbf{P},\mathbf{C}) &= \mathrm{H}(\mathbf{P} \mid (\mathbf{K},\mathbf{C})) + \mathrm{H}(\mathbf{K},\mathbf{C}) = \mathrm{H}(\mathbf{K} \mid \mathbf{C}) + \mathrm{H}(\mathbf{C}).\end{aligned}$$

□

Mějme šifrový text „WNAJW“, o kterém víme, že vznikl zašifrováním anglického slova posuvnou šifrou. Projdeme-li všech 26 klíčů, zjistíme, že existují pouze dvě anglická slova, ze kterých mohl šifrový text vzniknout. Jsou to slovo „river“ (posunem každého znaku o 5 písmen v abecedě) a slovo „arena“ (posunem o 22). Ostatní klíče vedou k otevřenému textu, který nedává smysl. To, že se nám podařilo zredukovat množinu 26 možných klíčů na 2 klíče, je tedy dáno tím, že lidský jazyk je zatížen velkou redundancí.

Mějme šifru $(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$ a náhodnou veličinu $\mathbf{P}^n$, která odpovídá volbě n po sobě jdoucích otevřených textů. Pro každé $y \in \mathcal{C}^n$ definujme množinu všech možných klíčů

$$\mathcal{K}(y) = \{\, k \in \mathcal{K} : \Pr(\mathbf{P}^n{=}D(k,y)) > 0 \,\}.$$

Ve shora uvedeném případě máme $\mathcal{K}(\text{„WNAJW“}) = \{5, 22\}$. Čím více šifrového textu máme k dispozici, tím víc klíčů budeme schopni vyloučit. Přirozená otázka je: kolik šifrového textu je v průměru třeba k tomu, abychom mohli určit klíč jednoznačně? Než na tuto otázku odpovíme, budeme muset kvantifikovat pojem redundance jazyka. Začneme tím, že zjistíme, jak velké množství informace je v průměru obsaženo v jednom písmeně jazyka. Tuto kvantitu nazýváme entropie jazyka. V případě jazyka, který má nulovou redundanci, je množství informace v jednom písmeně zřejmě $\log_2 26 \approx 4{,}70$. Na písmena anglického jazyka bychom se mohli v prvním přiblížení dívat jako na náhodnou veličinu $\mathbf{P}$ s rozdělením pravděpodobnosti, které odpovídá frekvenci jednotlivých písmen v jazyce. V takovém případě dostáváme entropii $\mathrm{H}(\mathbf{P}) \approx 4{,}19$. Ale jazyk není jen náhodná posloupnost písmen s určitým rozdělením. Každé písmeno je zasazeno do určitého kontextu. Vynecháme-li některá písmena v textu, můžeme být stále schopni zjistit původní význam z okolních písmen. Víme například, že některé bigramy se v jazyce vyskytují častěji než jiné. Označíme-li náhodnou veličinu $\mathbf{P}^2$, jejíž rozdělení odpovídá frekvenci bigramů v jazyce, pak $\mathrm{H}(\mathbf{P}^2)/2$ udává průměrné množství informace obsažené v jednom písmenu bigramu. Pro anglický jazyk máme $\mathrm{H}(\mathbf{P}^2)/2 = 3{,}9$. Takto bychom mohli pokračovat dál a počítat $\mathrm{H}(\mathbf{P}^n)/n$ jako průměrné množství informace v jednom písmenu n-gramu. Tato hodnota bude klesat s rostoucím n, protože s rozšiřujícím se kontextem klesá informační význam písmene.

Definice. Nechť L je jazyk a $\mathbf{P}^n$ je náhodná veličina, jejíž rozdělení odpovídá n-gramům v jazyce L, potom

$$\mathrm{H}_L = \lim_{n\to\infty} \frac{\mathrm{H}(\mathbf{P}^n)}{n}$$

nazýváme *entropie* jazyka L a

$$R_L = 1 - \frac{\mathrm{H}_L}{\log_2|\mathcal{P}|}$$

nazýváme *redundance* jazyka L.

Experimentální výsledky ukazují, že v případě anglického jazyka máme $1{,}0 \le \mathrm{H}_L \le 1{,}5$. To znamená, že každé písmeno anglického textu nese pouze $\mathrm{H}_L \approx 1{,}25$ bitů informace. Redundance potom vychází $0{,}68 \le R_L \le 0{,}79$, což znamená, že zhruba 75 % znaků v anglickém textu je z informačního hlediska nadbytečných. Přesněji řečeno, pro dostatečně velké n existuje kódování n-gramů, kterým lze docílit o 75 % kratšího textu.

Věta 4.2. *Nechť $(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$ je šifra a nechť R_L je redundance jazyka otevřených textů. Jsou-li klíče voleny z $\mathcal{K}$ náhodně s rovnoměrným rozdělením, pak střední hodnota počtu všech možných klíčů pro šifrový text délky n je*

$$\mathrm{E}|\mathcal{K}(\mathbf{C}^n)| \ge \frac{|\mathcal{K}| \cdot |\mathcal{P}|^{n(1-R_L)}}{|\mathcal{C}|^n}.$$

Důkaz. Podle tvrzení 4.1 máme

$$\mathrm{H}(\mathbf{K} \mid \mathbf{C}^n) = \mathrm{H}(\mathbf{K}) + \mathrm{H}(\mathbf{P}^n) - \mathrm{H}(\mathbf{C}^n). \tag{4.1}$$

Vzhledem k tomu, že $\mathrm{H}(\mathbf{P}^n)/n$ je nerostoucí posloupnost, máme

$$\mathrm{H}(\mathbf{P}^n) \ge n\,\mathrm{H}_L = n(1 - R_L)\log_2|\mathcal{P}|.$$

Podle věty o maximální entropii je

$$\mathrm{H}(\mathbf{C}^n) \le \log_2|\mathcal{C}^n| = n\log_2|\mathcal{C}|.$$

Dosazením těchto dvou poznatků do rovnice (4.1) dostáváme

$$\mathrm{H}(\mathbf{K} \mid \mathbf{C}^n) \geq \mathrm{H}(\mathbf{K}) + n(1 - R_L)\log_2|\mathcal{P}| - n\log_2|\mathcal{C}|. \tag{4.2}$$

Nyní rozepíšeme $\mathrm{H}(\mathbf{K} \mid \mathbf{C}^n)$ podle definice, provedeme horní odhad podle věty o maximální entropii a následně pomocí Jensenovy nerovnosti.

$$\begin{aligned}\mathrm{H}(\mathbf{K} \mid \mathbf{C}^n) &= \sum_{y \in \mathcal{C}^n} \Pr(\mathbf{C}^n{=}y)\,\mathrm{H}(\mathbf{K} \mid \mathbf{C}^n{=}y) \leq \sum_{y \in \mathcal{C}^n} \Pr(\mathbf{C}^n{=}y)\log_2|\mathcal{K}(y)| \\ &\leq \log_2 \sum_{y \in \mathcal{C}^n} \Pr(\mathbf{C}^n{=}y)|\mathcal{K}(y)| = \log_2 \mathrm{E}|\mathcal{K}(\mathbf{C}^n)|.\end{aligned} \tag{4.3}$$

Složením odhadů (4.2) a (4.3) dostáváme

$$\log_2 \mathrm{E}|\mathcal{K}(\mathbf{C}^n)| \geq \mathrm{H}(\mathbf{K}) + n(1 - R_L)\log_2|\mathcal{P}| - n\log_2|\mathcal{C}|.$$

K dokončení důkazu stačí dodat, že $\mathrm{H}(\mathbf{K}) = \log_2|\mathcal{K}|$, protože klíče jsou voleny náhodně s rovnoměrným rozdělením. □

Přirozenou otázkou je, jaké minimální množství šifrového textu je třeba k tomu, aby na danou šifru bylo možné úspěšně provést known-ciphertext attack. V tuto chvíli nás nezajímá náročnost takového útoku, ale pouze to, zda při něm lze teoreticky uspět. Jinými slovy, předpokládáme, že výpočetní možnosti útočníka jsou neomezené, a ten je tedy schopen provést útok hrubou silou. Pro šifrový text y lze tento útok úspěšně provést, když množina možných klíčů $\mathcal{K}(y)$ je jednoprvková. Položíme-li v předchozí větě $\mathrm{E}|\mathcal{K}(\mathbf{C}^n)| = 1$, pak dostáváme následující nerovnost pro n:

$$n \geq \frac{\log_2|\mathcal{K}|}{\log_2|\mathcal{C}| - (1 - R_L)\log_2|\mathcal{P}|}.$$

Výraz na pravé straně nerovnosti nazýváme *vzdálenost jednoznačnosti* šifry. V případě, že $|\mathcal{C}| = |\mathcal{P}|$, se nerovnost zjednoduší na

$$n \geq \frac{\log|\mathcal{K}|}{R_L \log|\mathcal{P}|}.$$

Při použití substituční šifry na anglickém textu máme $|\mathcal{K}| = 26!$, $|\mathcal{P}| = |\mathcal{C}| = 26$ a $R_L = 0{,}75$. Po dosazení do vzorce zjistíme, že k jednoznačnému dešifrování může stačit i šifrový text o délce pouhých 26 znaků.

Při použití Vigenèrovy šifry s klíčem délky k na anglickém textu máme $|\mathcal{K}| = 26^k$, $|\mathcal{P}| = |\mathcal{C}| = 26^k$ a $R_L = 0{,}75$. Po dosazení do vzorce zjistíme, že k jednoznačnému dešifrování je třeba šifrový text o délce alespoň $\frac{4}{3}$ bloků, čili $\frac{4}{3}k$ znaků. Tento odhad je zjevně nutno brát s jistou rezervou. Zvažme, jak by probíhal útok, máme-li k dispozici šifrový text délky $\frac{4}{3}k$ znaků. Vzhledem k povaze Vigenèrovy šifry je na místě rozdělit šifrový text na čtyři úseky $y_1, \ldots, y_4$, každý délky $\frac{1}{3}k$. Úseky y_1 a y_4 jsou šifrovány stejným klíčem zatímco úseky y_2 a y_3 jsou šifrovány klíčem, který je náhodný a nikde se neopakuje. Odečteme-li y_1 od y_4, eliminujeme tím klíč a získáme rozdíl příslušných částí otevřeného textu. Z tohoto rozdílu pak můžeme být schopni rekonstruovat úseky x_1 a x_4 otevřeného textu. Známe-li tyto, pak využijeme redundanci jazyka a domyslíme si x_2 a x_3, které není možné zjistit žádným jiným způsobem. Všimněme si, že zde využíváme redundanci jazyka velmi nerovnoměrným způsobem, protože doplňujeme celé souvislé chybějící úseky textu. Naproti tomu hodnota R_L sleduje průměrnou redundanci jazyka a nereflektuje tedy danou situaci. Skutečná vzdálenost jednoznačnosti bude blíže hodnotě $2k$. Máme-li šifrový text délky $2k$, pak jej můžeme rozdělit na dva úseky stejné délky, z nichž oba jsou šifrovány stejným klíčem, a postupovat jako při rekonstrukci úseků y_1 a y_4 výše.

Při použití šifry AES s klíčem délky 128 bitů je vzdálenost jednoznačnosti závislá na tom, jakým způsobem zakódujeme otevřený text do bloku 16 bajtů, se kterými tato šifra pracuje. Nejjednodušší je kódovat jeden znak do každého bajtu, čímž dostáváme $\log_2|\mathcal{P}| = \log_2 26^{16} \approx 75$, a dále máme $\log_2|\mathcal{C}| = 128$ a $\log_2|\mathcal{K}| = 128$. K jednoznačnému dešifrování může stačit šifrový text o délce pouhých 1,17 bloků, tj. 19 znaků.

5. ABSOLUTNÍ BEZPEČNOST

V tomto oddílu pracujeme všude se šifrou $(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$. Náhodné veličiny, které odpovídají volbě klíče, volbě otevřeného textu a výslednému šifrovému textu, značíme $\mathbf{K}$, $\mathbf{P}$ a $\mathbf{C}$. Předpokládáme, že $\mathbf{P}$ a $\mathbf{K}$ jsou nezávislé.

Definice. Říkáme, že šifra $(\mathcal{P}, \mathcal{C}, \mathcal{K}, E, D)$ je *absolutně bezpečná*, jestliže pro všechna $x \in \mathcal{P}$ a $y \in \mathcal{C}$ taková, že $\Pr(\mathbf{C}{=}y) \neq 0$, platí $\Pr(\mathbf{P}{=}x \mid \mathbf{C}{=}y) = \Pr(\mathbf{P}{=}x)$.

Tvrzení 5.1. *Šifra je absolutně bezpečná právě tehdy, když* $\mathbf{P}$ *a* $\mathbf{C}$ *jsou nezávislé náhodné veličiny.*

Důkaz. Absolutní bezpečnost šifry je ekvivalentní podmínce, že

$$\Pr(\mathbf{P}{=}x) = \Pr(\mathbf{P}{=}x \mid \mathbf{C}{=}y) = \frac{\Pr(\mathbf{P}{=}x, \mathbf{C}{=}y)}{\Pr(\mathbf{C}{=}y)}$$

pro všechna x a y taková, že $\Pr(\mathbf{C}{=}y) \neq 0$. Tato podmínka je ekvivalentní podmínce, že $\mathbf{P}$ a $\mathbf{C}$ jsou nezávislé náhodné veličiny, stačí pouze uvážit, že kdykoliv $\Pr(\mathbf{C}{=}y) = 0$, pak také $\Pr(\mathbf{P}{=}x, \mathbf{C}{=}y) = 0$. □

Spojením předchozího tvrzení s důsledkem 3.6 dostáváme:

Důsledek 5.2. *Následující podmínky jsou ekvivalentní:*

1. *šifra je absolutně bezpečná;*
2. $\mathrm{H}(\mathbf{P} \mid \mathbf{C}) = \mathrm{H}(\mathbf{P})$*;*
3. $\mathrm{H}(\mathbf{C} \mid \mathbf{P}) = \mathrm{H}(\mathbf{C})$.

Věta 5.3. *Jestliže klíč je volen náhodně s rovnoměrným rozdělením, pak Vernamova šifra je absolutně bezpečná.*

Důkaz. Pro všechna $x \in \mathcal{P}$ a $y \in \mathcal{C}$ platí

$$\Pr(\mathbf{P}{=}x, \mathbf{C}{=}y) = \Pr(\mathbf{P}{=}x, \mathbf{K}{=}x \oplus y) = \Pr(\mathbf{P}{=}x)\Pr(\mathbf{K}{=}x \oplus y) = \Pr(\mathbf{P}{=}x)|\mathcal{K}|^{-1}.$$

Zde jsme využili nezávislost náhodných veličin $\mathbf{P}$ a $\mathbf{K}$ a rovnoměrného rozdělení klíčů. Podle důsledku 1.2 pak dostáváme

$$\Pr(\mathbf{C}{=}y) = \sum_{x \in \mathcal{P}} \Pr(\mathbf{P}{=}x, \mathbf{C}{=}y) = \sum_{x \in \mathcal{P}} |\mathcal{K}|^{-1} \Pr(\mathbf{P}{=}x) = |\mathcal{K}|^{-1}.$$

Spojením obou výsledků zjišťujeme, že

$$\Pr(\mathbf{P}{=}x \mid \mathbf{C}{=}y) = \frac{\Pr(\mathbf{P}{=}x, \mathbf{C}{=}y)}{\Pr(\mathbf{C}{=}y)} = \Pr(\mathbf{P}{=}x)$$

pro všechna $x \in \mathcal{P}$ a $y \in \mathcal{C}$, což je definice absolutní bezpečnosti. □

Lemma 5.4. *Platí* $\mathrm{H}(\mathbf{C}) \geq \mathrm{H}(\mathbf{P})$.

Důkaz. Podle důsledku 3.6 máme $\mathrm{H}(\mathbf{C} \mid \mathbf{K}) \leq \mathrm{H}(\mathbf{C})$. Znalost klíče nám dává jednoznačnou korespondenci mezi otevřenými a šifrovými texty, proto $\mathrm{H}(\mathbf{C} \mid \mathbf{K}) = \mathrm{H}(\mathbf{P} \mid \mathbf{K})$. Vzhledem k tomu, že veličiny $\mathbf{P}$ a $\mathbf{K}$ jsou nezávislé, podle důsledku 3.6 máme $\mathrm{H}(\mathbf{P} \mid \mathbf{K}) = \mathrm{H}(\mathbf{P})$. □

Věta 5.5 (Shannon)**.** *Jestliže je šifra absolutně bezpečná, pak* $\mathrm{H}(\mathbf{K}) \geq \mathrm{H}(\mathbf{P})$.

Důkaz. Všimněme si, že platí

$$\mathrm{H}(\mathbf{K}) + \mathrm{H}(\mathbf{P}) = \mathrm{H}(\mathbf{K}, \mathbf{P}) = \mathrm{H}(\mathbf{K}, \mathbf{P}, \mathbf{C}) \geq \mathrm{H}(\mathbf{P}, \mathbf{C}).$$

První rovnost plyne z věty 3.4 a předpokladu, že náhodné veličiny $\mathbf{K}$ a $\mathbf{P}$ jsou nezávislé. Druhá rovnost plyne z toho, že známe-li otevřený text a klíč, pak šifrový text je určen jednoznačně. Poslední nerovnost je z důsledku 3.7.

Spojením shora uvedené nerovnosti s tvrzením 3.5, důsledkem 5.2 a lemmatem 5.4 dostáváme $\mathrm{H}(\mathbf{K}) \geq \mathrm{H}(\mathbf{P}, \mathbf{C}) - \mathrm{H}(\mathbf{P}) = \mathrm{H}(\mathbf{C} \mid \mathbf{P}) = \mathrm{H}(\mathbf{C}) \geq \mathrm{H}(\mathbf{P})$. □

Má-li $\mathbf{P}$ rovnoměrné rozdělení, pak pro Vernamovu šifru máme $\mathrm{H}(\mathbf{K}) = \mathrm{H}(\mathbf{P})$. Z Shannonovy věty tedy plyne, že v případě, že $\mathbf{P}$ má rovnoměrné rozdělení, pak žádná absolutně bezpečná šifra není z hlediska délky klíče efektivnější než Vernamova šifra.

6. DALŠÍ TYPY BEZPEČNOSTI

Jak jsme v závěru předchozího oddílu právě zjistili, absolutně bezpečné šifry jsou velmi nepraktické, protože vyžadují, aby délka klíče, který je třeba sdělit utajeným způsobem, byla alespoň tak velká, jako délka šifrované zprávy. V praxi proto uvažujeme jiné druhy bezpečnosti.

Říkáme, že kryptografický systém je *výpočetně bezpečný*, jestliže nejlepší známý útok, kterým ho lze prolomit, vyžaduje alespoň N operací, kde N je velké číslo převyšující výpočetní možnosti útočníka. V současné době lze pro běžné účely považovat za výpočetně bezpečné $N \geq 2^{85}$.

Říkáme, že kryptografický systém je *dokazatelně bezpečný*, jestliže jeho prolomením by bylo možné současně řešit nějaký dobře zkoumaný matematický problém, pro který není známa efektivní metoda řešení. Může se jednat například o problém faktorizace nebo diskrétního logaritmu. Jedná se o stejný princip, jako v případě dokazování NP-úplnosti nějakého problému.

LITERATURA

[1] Shannon, C.: A mathematical theory of communication. *Bell System Technical Journal, The*, ročník 27, č. 3, July 1948: s. 379–423, ISSN 0005-8580, doi:10.1002/j.1538-7305.1948.tb01338.x.
URL `http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x`