1 Teorie informace

Mějme nějakou abecedu $\mathbb{A}$ nad konečným počtem symbolů $a_1, \ldots, a_n$. Při používání $\mathbb{A}$ ať se každé $a_i \in \mathbb{A}$ vyskytuje s frekvencí $p_{a_i} = p_i$. Je tedy $\sum p_i = 1$.

Předpokládejme, že bychom chtěli zavést jakýsi binární těsnopis, který by komunikaci pomocí symbolů abecedy $\mathbb{A}$ co nejvíce komprimoval. Jde o jednoduchou myšlenku, kterou ilustrujeme na příkladě, ve kterém frekvence p_i upravíme tak, aby souvislost s binárními zápisem co nejlépe vynikla.

Za symboly abecedy $\mathbb{A}$ budeme považovat smutně realistické frekvence povelů psovi Hafíkovi uvedené v následující tabulce.

Symbol	Povel	frekvence	binárně
a_1	SEDNI	1/4	00
a_2	LEHNI	1/4	01
a_3	DEJ POKOJ	1/4	10
a_4	DO BOUDY	1/8	110
a_5	APORT	1/16	1110
a_6	HLEDEJ	1/32	11110
a_7	PAC	1/64	111110
a_8	VEM SI HO	1/64	111111

Nejméně frekventované povely jsou nejpřekvapivější, a proto lze říci, že nesou nejvíce informace. To je vyjádřeno i tím, že při snaze o co nejúspornější záznam komunikace je jim věnováno nejvíce bitů. Jejich počet budeme považovat za (binární) *informační obsah* povelu. V našem případě to jsou po řadě čísla 2, 2, 2, 3, 4, 5, 6, 6. Obecně nemusíme dostat celé číslo – hodnotou informačního obsahu $I(a_i)$ symbolu a_i definujeme jako $\lg(1/p_i)$, kde lg znamená logaritmus při základu dva.

Podobný postup jako ten, který jsme naznačili výše, se používá při některých typech komprimace textu. Podle frekvence výskytu slov či frází se vybuduje binární strom, na jehož koncových uzlech jsou umístěna slova či fráze tak, aby součet frekvencí koncových uzlů levé části byl zhruba stejný jako součet frekvencí koncových uzlů v pravé části (oba součty tedy aproximují hodnotu 1/2). Rekurzivně se pak pokračuje i při dalších větveních. Takto vytvořeným kódům se říká *Huffmanovy*.

Hodnoty p_i jsme volili jako pravděpodobnosti výskytu symbolu a_i, tedy $p_i = \Pr(a_i)$. Pro další úvahy učiníme zjednodušující předpoklad, že sled symbolů abecedy $\mathbb{A}$ není kontextově závislý. To je samozřejmě předpoklad, který ve většině aplikací splněn nebude, nicméně bez něj by se následující úvahy nesmírně komplikovaly. Tyto úvahy přitom vedou k silné teorii, která se zpětně využívá i tak, že se pořadí vysílaných symbolů na určitém dostatečně velkém úseku permutuje, takže jejich pořadí se skutečně pak jako kontextově nezávislé jeví.

Jinými slovy, budeme předpokládat, že pro slovo $a_{i_1} \dots a_{i_k} \in \mathbb{A}^*$ máme

$$\Pr(a_{i_1} \dots a_{i_k}) = \Pr(a_{i_1}) \dots \Pr(a_{i_k}) = p_{i_1} \dots p_{i_k} = \prod p_{i_j}.$$

Z kontextové nezávislosti vyplývá, že z předchozího toku symbolů nelze odvodit obsah následující. Proto se míra informace obsažená ve dvou za sebou následujících úsecích rovná součtu informačního obsahů jednotlivých úseků. Dostáváme

$$I(a_{i_1} \dots a_{i_k}) = I(a_{i_1}) + \dots + I(a_{i_k}) = \sum \lg \frac{1}{p_{i_j}} = \lg \frac{1}{\prod p_{i_j}}.$$

Slovo „entropie“ je používáno v mnoha různých významech. Obecně znamená absenci nějakého řádu, který by umožnil rozlišovat mezi jednotlivými součástmi systému. Etymologicky „entropický“ znamená „dovnitř obrácený“, což vzniklo jako snaha popsat v termodynamice tu část energie systému, kterou nelze použít pro vykonání práce. V naší situaci se absencí řádu možná poněkud překvapivě míní malá odlišnost frekvencí symbolů. Jde o pojem definovaný, kde vést úvahy o vhodnosti volby slova má jen omezený význam. Poznamenejme však, že jazyk, ve kterém by všechna slova měla stejnou frekvenci, by byl pro zvládnutí skutečně velmi obtížný.

Stanovení frekvencí p_i vyžaduje nějaký tok dat symbolů abecedy $\mathbb{A}$. Mluvíme o *informačním zdroji* $\mathcal{A}$ – formálně ho lze například popsat jako nekonečnou posloupnost symbolů abecedy $\mathbb{A}$, přičemž frekvence p_i vyjadřují limitní hodnoty počítané z konečných počátečních úseků. *Entropií* informačního zdroje $\mathcal{A}$ se pak rozumí průměrná informační hodnota symbolů abecedy $\mathbb{A}$. Používáme označení $\mathcal{H}(\mathcal{A})$, takže platí

$$\mathcal{H}(\mathcal{A}) = \sum p_i I(a_i) = \sum p_i \lg(\frac{1}{p_i}).$$

Pro náš úvodní příklad vychází entropie $3 \cdot 2/4 + 3/8 + 4/16 + 5/32 + 2 \cdot 6/64 = 79/32$. Pokud by všech osm povelů mělo stejnou frekvenci, dostali bychom entropii vyšší, a to $8 \cdot 3/8 = 3$. Prvá souvislost s entropickou funkcí H se ukazuje v případě, kdy informační zdroj $\mathcal{A}$ je binární (má tedy pouze dva symboly). Pokud $p = p_1$, tak $p_2 = 1 - p$ a $\mathcal{H}(\mathcal{A}) = p \lg p^{-1} + (1-p) \lg(1-p)^{-1} = H(p)$. Tato souvislost však nebude tou hlavní, k té dospějeme teprve úvahami o spolehlivosti kanálu.

Entropii informačního zdroje můžeme neformálně chápat jako průměrnou sdělnost znaku abecedy $\mathbb{A}$, což například znamená, že při převodu do optimalizovaného binárního zápisu je na záznam zprávy délky k potřeba $k \cdot \mathcal{H}(\mathcal{A})$ bitů (ve skutečnosti to bude o něco více díky necelým hodnotám). Na informační obsah symbolu se též můžeme dívat jako na míru překvapení, že se ve zprávě symbol objevil (tak jako pes Hafík bude spíše překvapen, že si má někoho vzít než že má dát pokoj). Entropii tedy můžeme také chápat jako průměrnou míru překvapení.

Předpokládejme nyní, že informační zdroj $\mathcal{A}$ prochází nějakou transformací (obvykle se mluví o průchodu kanálem), po jejímž provedení se mění na informační zdroj $\mathcal{B}$. Symboly $\mathcal{A}$ nechť se vyskytují s frekvencemi p_i a synboly $\mathcal{B}$ s

frekvencemi q_j. Předpokládáme, že kanál mění symbol na symbol, a to tak, že i-tý vstupní symbol a_i přejde na j-tý výstupní symbol b_j s pravděpodobností p_{ij}. Počet vstupních symbolů buď roven n a počet výstupních symbolů ať je m. Z pravděpodobností p_{ij} vytvořme matici $P = (p_{ij})$ rozměrů $n \times m$. Vidíme, že platí

$$(p_1, \ldots, p_n)P = (q_1, \ldots, q_m).$$

Formálně vzato můžeme transformaci $\mathcal{A} \to \mathcal{B}$ považovat za reversibilní, tedy uvažovat opačný průchod kanálem. Pravděpodobnost přechodu b_j na a_i označíme q_{ji}. Hodnoty p_{ij} a q_{ji} jsou vlastně podmíněné pravděpodobnosti $\Pr(b_j|a_i)$ a $\Pr(a_i|b_j)$. Ze vztahu

$$\Pr(a_i)\ \Pr(b_j|a_i) = \Pr(a_i, b_j) = \Pr(b_j)\ \Pr(a_i|b_j)$$
$$\text{vyplývá}\ \ p_i p_{ij} = q_j q_{ji},\ \text{takže}\ \ q_{ji} = \frac{p_i}{q_j} p_{ij}.$$

Uvažme nyní informační obsah zjištění při přijetí symbolu b_j, že byl vyslán symbol a_i. Frekvence jsou v takto zúženém případě rovny $q_{j1}, \ldots, q_{jn}$, takže relativní informační obsah $I(a_i|b_j)$ je roven $\lg(1/q_{ji})$. Můžeme ho považovat za míru překvapení, když se dozvíme, že byl vyslán symbol a_i. Pokud by obě abecedy splývaly a informační kanál byl pouze identickým opakováním, tak bude $q_{jj} = 1$ a $q_{ji} = 0$ pro $i \neq j$. V takovém případě je překvapení při přijetí shodného symbolu nulové, zatímco při přijetí odlišného symbolu (k čemuž nemůže dojít) by bylo překvapení nekonečně velké. Pokud známe pouze informační zdroj $\mathcal{B}$, je pro nás informace odpovídající zjištění vyslaného symbolu a_i nedostupná. Vážený průměr $\mathcal{H}(\mathcal{A}|b_j)$ informačních obsahů $I(a_i|b_j)$ je tedy roven průměrnému ztracenému informačnímu obsahu při příjmu symbolu b_j, a vážený průměr $\mathcal{H}(\mathcal{A}|\mathcal{B})$ přes všechna b_j udává průměrnou ztrátu informačního obsahu na jeden symbol. Máme

$$\mathcal{H}(\mathcal{A}|b_j) = \sum_i q_{ji} \lg \frac{1}{q_{ji}} \quad \text{a} \quad \mathcal{H}(\mathcal{A}|\mathcal{B}) = \sum_j q_j \mathcal{H}(\mathcal{A}|b_j) = \sum_j q_j \sum_i q_{ji} \lg \frac{1}{q_{ji}},$$

Průměrný informační obsah jednoho symbolu z $\mathcal{A}$ je rovna $\mathcal{H}(\mathcal{A})$, takže

$$I(\mathcal{A}, \mathcal{B}) = \mathcal{H}(\mathcal{A}) - \mathcal{H}(\mathcal{A}|\mathcal{B})$$

udává průměrný informační obsah přijatého symbolu. Je ovšem otázka, zda tento obsah lze skutečně efektivně zpřístupnit. Ukážeme na jednom konkrétním případu kanálu, že pomocí samoopravných kódů to možné je.

Typem kanálu, který budeme uvažovat, je nejjednodušší netriviální případ, a to takzvaný *binární symetrický kanál*, zkráceně BSC, ve kterém se vysílají i přijímají dva symboly, přičemž k chybě dochází s pravděpodobností $p = p_{12} = p_{21}$. Předpokládáme, že oba vysílané symboly jsou stejně frekventované, tedy $p_1 = p_2 = 1/2$. Máme $p_{11} = p_{22} = 1 - p$ a

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

takže $q_1 = q_2 = 1/2$, $q_{12} = q_{21} = p$ a $q_{11} = q_{22} = 1 - p$.

Vidíme, že $\mathcal{H}(\mathcal{A}|\mathcal{B}) = 2(1/2)(p\lg(1/p) + (1-p)\lg(1/(1-p))) = H(p)$. Dále $\mathcal{H}(\mathcal{A}) = 2(1/2)H(1/2) = 1$, takže $I(\mathcal{A},\mathcal{B}) = 1 - H(p)$. Hodnota $C(p) = 1 - H(p)$ se nazývá *kapacita kanálu*. Pravděpodobnosti p se říká *chybovost* kanálu.

Nyní uvedeme (hlavní) Shannonovu větu, zvanou též Základní věta teorie informací. Připomeňme, že informačním poměrem (nosností) binárního kódu C délky n, který má k kódových slov, rozumíme hodnotu $\alpha(C) = (\lg k)/n$. Dekódováním metodou nejbližšího slova rozumíme algoritmus, ve kterém přijatému slovu v přiřadíme kódové slovo u, které má nejmenší Hammingovu vzdálenost $d(u,v)$, a to jen tehdy, pokud takové kódové slovo u existuje.

Věta 1.1. *Ať je dán BSC kanál s chybovostí p, kde $0 < p < 1/2$. Pak pro všechna $\varepsilon > 0$ a $\delta > 0$ existuje $n_0 > 1$, že pro každé $n > n_0$ existuje binární kód C délky n takový, že $1-H(p) > \alpha(C) \geq 1-H(p)-\varepsilon$, ve kterém dekódování metodou nejbližšího kódového slova selhává s pravděpodobností menší než δ.*

Nástin důkazu. Podrobný a přesný důkaz uvedeme v závěru kapitoly. Zde je cílem seznámení s jeho hlavními myšlenkami. Přitom se nebudeme vyjadřovat vždy formálně zcela přesně. Ukážeme, že našemu požadavku vyhoví každý kód C, který je vybrán dostatečně náhodně a který má pro nějaké $R < 1 - H(p)$ přesně 2^{nR} slov. Je tedy $R = \alpha(C)$. Přitom R volíme při zadaném ε tak, aby existovalo nějaké $\gamma > 0$, že $1 - H(p) - \varepsilon \leq R \leq 1 - H(p) - \gamma$. Náhodnost C znamená, že kódových slov v kouli o poloměru j je relativně stejně jako v celém kódu. Čili předpokládáme, že

$$|\{u \in C;\ d(u,v) \leq j\}| = \frac{2^{nR}}{2^n} \sum_{i \leq j} \binom{n}{i},$$

pro všechna j, která jsou dostatečně malá. Na pravé straně rovnosti nemusí být samozřejmě celé číslo; ve skutečnosti nejde o rovnost, ale o průměrný počet kódových slov. Náhodnost C zaručuje, že opakované náhodné volby koule o poloměru j dávají tento průměrný počet.

Pro dostatečně velká n lze předpokládat, že chyb při přenosu je asi pn. Přesněji to vymezuje zákon velkých čísel, ze kterého, jak níže uvidíme, vyplyne, že přijaté slovo v se od vyslaného u liší natolik málo, že lze předpokládat, že $d(u,v)$ je v blízkosti pn. Hodnotu blízkou pn budeme volit jako poloměr koule, kterou použijeme pro dekódování. Problém může nastat v tom, že by vedle u existovalo ještě nějaké $u' \in C$, které by také leželo ve vzdálenosti málo odlišné od pn, případně ve vzdálenosti menší. Ukážeme, že z pravděpodobnostního hlediska tomu tak není. Vtip je v tom, že existenci u máme zaručenou přenosem, ale existence u' už je záležitostí náhodnou, na kterou lze použít předpoklad o náhodném charakteru kódu C. V kouli o poloměru pn najdeme průměrně kódových slov

$$\frac{2^{nR}}{2^n} \sum_{i \leq pn} \binom{n}{i} \leq 2^{nH(p)} \frac{2^{nR}}{2^n} = 2^{-n(1-H(p)-R)} \leq 2^{-n\gamma}.$$

Uvedená nerovnost plyne z Lemmatu **??**. Vidíme, že s rostoucím n se pravděpodobnost výskytu jakéhokoliv slova v kouli poloměru pn blíží nule. Proto se blíží nule i pravděpodobnost existence kódového slova $u' \neq u$. □

Jako fakt nyní uvedeme takzvaný Slabý zákon velkých čísel. Ten se dokazuje v teorii pravděpodobnosti. Je založen na pradávné zkušenosti z oblasti hazardních her, že při dostatečném počtu opakování hodů se dosahuje stejného stabilního průměrného chování. Dá se s jistou nadsázkou také říci, že teorie pravděpodobnosti byla vybudována právě tak, aby v ní tento zákon platil. Proto není nutné na jeho důkaz formálně odkazovat.

Věta 1.2. *Ať $x_1, x_2, \ldots$ je posloupnost, jejíž členy nabývají náhodně reálných hodnot $a_1, a_2, \ldots, a_m$, a to s pravděpodobnostmi $p_1, \ldots, p_m$. Položme $\mu = \sum p_j a_j$. Pak pro každé $\varepsilon > 0$ je*

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right| > \varepsilon\right) = 0.$$

Poznamenejme, že členům x_i (jednotlivým „tahům“) se v teorii pravděpodobnosti říká náhodné proměnné.

Při přenosu pomocí BSC s chybovostí p odpovídají proměnné x_i úspěšnosti přenosu i-tého symbolu. Nedošlo-li k chybě, položíme $x_i = 0$, naopak při chybě bude $x_i = 1$. Je-li $u = u_1 \ldots u_n$ vyslané slovo a $v = v_1 \ldots v_n$ slovo přijaté, tak $\sum x_i = d(u, v)$. Podle Věty 1.2 je

$$\lim_{n\to\infty} \Pr\left(|d(u, v) - pn| > n\varepsilon\right) = 0.$$

Podmínku $|d(u, v) - pn| \leq n\varepsilon$ můžeme také zapsat jako $v \in M(u, \varepsilon)$, kde $M(u, \varepsilon) = \{x \in \mathbb{F}_2^n;\ n(p - \varepsilon) \leq d(u, x) \leq n(p + \varepsilon)\}$. Zákon velkých čísel tedy říká, že pro libovolné $\varepsilon > 0$ lze pro stanovené $\delta > 0$ najít n_0 tak, že pro každé $n > n_0$ je $v \in M(u, \varepsilon)$ s pravděpodobností větší než $1 - \delta$. Tato skutečnost doplňuje nástin důkazu Věty 1.1 a níže ji ještě několikrát využijeme.

Pravděpodobnost, že $u = u_1 \ldots u_n$ je přeneseno jako $v = v_1 \ldots v_n$ lze také vyjádřit jako $p^d(1 - p)^{n-d}$, kde $d = d(u, v)$. Předokládáme $0 < p < 1/2$, takže tato pravděpodobnost s rostoucí vzdáleností od u klesá. Prvek $v \in M(u, \varepsilon)$ má nejmenší možnou vzdálenost od u rovnou $n(p - \varepsilon)$, a proto pravděpodobnost transformace u na $v \in M(u, \varepsilon)$ je shora omezena hodnotou

$$p^{n(p-\varepsilon)}(1 - p)^{n(1-p+\varepsilon)} = p^{np}(1 - p)^{n(1-p)} \left(\frac{1 - p}{p}\right)^{n\varepsilon}.$$

Protože $p^{np}(1-p)^{n(1-p)} = 2^{np \lg p} 2^{n(1-p)\lg(1-p)} = 2^{-nH(p)}$, tak můžeme vyslovit

Lemma 1.3. *Při přenosu BSC kanálem s chybovostí $0 < p < 1/2$ uvažme binární slova u a v délky $n \geq 1$. Ať $0 < \varepsilon < 1$. Je-li u slovo vyslané a $v \in M(u, \varepsilon)$, tak pravděpodobnost toho, že v bude slovo přijaté, je shora omezena hodnotou $2^{-nH(p)}((1 - p)/p)^{n\varepsilon}$.*

Kapacita kanálu je ostrým předělem pro nosnost (informační poměr) $\alpha(C)$ kódu C. Podle Shannonovy věty 1.1 lze konstruoovat libovolně spolehlivé kódy pokud připustíme, aby nosnost byla menší než kapacita kanálu. Přitom není potřeba vymýšlet nějaké chytré strategie dekódování – vystačíme s dekódováním

pomocí nejbližšího slova. Smyslem obrácené Shannonovy věty 1.5 bude naopak dokázat, že při použití libovolné metody dekódování nemůžeme dosáhnout uspokojivé spolehlivosti přenosu, pokud nosnost kódu převýší kapacitu kanálu. Tato skutečnost odpovídá pozorováním o ztrátě informace, ukážeme navíc, že pro dostatečně velká n je kvalita přenosu „libovolně špatná“.

Spolehlivost dekódování se vždy vztahuje k nějakému algoritmu, který z přijatého slova vytváří slovo kódové. Tento algoritmus musí být efektivně dostupný, pro naše účely však stačí, abychom si uvědomili, že musí realizovat nějaké zobrazení $D : \mathbb{F}_2^n \to C$, kde n je délka binárního kódu C. *Spolehlivostí dekódování* D se rozumí průměrná pravděpodobnost, že $D(v) = u$, kde (u, v) probíhá všechny možné dvojice odpovídající vyslanému a přijatému slovu. Je to ovšem průměr vážený, kde váha dvojice (u, v) je dána pravděpodobností, že v bude slovo přijaté. Tato pravděpodobnost závisí na chybovosti kanálu p (a už jsme se jí výše zabývali). *Spolehlivostí kódu* C pak budeme rozumět maximální spolehlivost počítanou přes všechna zobrazení $D : \mathbb{F}_2^n \to C$. Zdůrazněme, že spolehlivost se počítá vždy relativně vůči dané chybovosti.

Při důkazu Věty 1.5 se bude hodit následující elementární pozorování.

Lemma 1.4. *Jestliže jev A nastává s pravděpodobností a a jev B s pravděpodobností b, tak jev $A \cap B$ nastává s pravděpodobností alespoň $a + b - 1$.*

Důkaz. Označme x pravděpodobnost jevu $A \cap B$. Potom $a - x$ je pravděpodobnost jevu $A \setminus B$. Ta je $\leq$ pravděpodobnost doplňkového jevu B', takže $a - x \leq 1 - b$. □

Věta 1.5. *Ať γ a p jsou reálná čísla, která splňují $0 < \gamma < 1$ a $0 < p < 1/2$. Uvažujme BSC chybovosti p. Pro každé $R > 1 - H(p)$ existuje celé číslo n_0, že každý binární kód délky $n \geq n_0$, který splňuje $\alpha(C) \geq R$, má spolehlivost $\leq \gamma$.*

Důkaz. Budeme postupovat sporem, tedy budeme předpokládat, že existuje $1 > \gamma > 0$ takové, že pro libovolně velká n lze nalézt kód C a zobrazení $D : \mathbb{F}_2^n \to C$ tak, aby průměrná (vážená) pravděpodobnost toho, že pro dvojici (u, v) vyslaného a přijatého slova platí $D(v) = u$, byla $\geq \gamma$.

Podle Zákona velkých čísel pro každé $\varepsilon > 0$ platí, že pro dostatečně velká n je $v \in M(u, \varepsilon)$ s pravděpodobností $\geq 1 - \gamma/2$. Podle Lemmatu 1.4 je pravděpodobnost souběhu jevů $D(v) = u$ a $v \in M(u, \varepsilon)$ alespoň $\gamma + (1 - \gamma/2) - 1 = \gamma/2$. Tedy

$$\Pr\left(v \in D^{-1}(u) \cap M(u, \varepsilon)\right) \geq \gamma/2.$$

Podle Lemmatu 1.3 je uvedená pravděpodobnost pro jedno konkrétní $u \in C$ shora omezena hodnotou

$$2^{-nH(p)} \left(\frac{1-p}{p}\right)^{n\varepsilon} \left|D^{-1}(u) \cap M(u, \varepsilon)\right|.$$

Součet těchto hodnot počítaný přes všechna $u \in C$ je tedy horním odhadem $\Pr\left(v \in D^{-1}(u) \cap M(u, \varepsilon)\right)$. Označme k počet kódových slov kódu C a

všimněme si, že množiny $D^{-1}(u)$, kde u probíhá C, jsou po dvou disjunktní. Dokázali jsme, že

$$2^n \geq \sum_{u \in C} \left| D^{-1}(u) \cap M(u, \varepsilon) \right| \geq \frac{k\gamma}{2} 2^{nH(p)} \left(\frac{p}{1-p} \right)^{n\varepsilon}.$$

Úvahu dovedeme ke sporu tím, že ukážeme, že při dostatečně malém $\varepsilon > 0$ a při dostatečně velkém n je pravá strana větší než 2^n. Podle předpokladu o $\alpha(C)$ je $\lg k \geq Rn$ a $R > 1 - H(p)$. Logaritmus pravé strany je tedy roven

$$Rn + nH(p) + \lg \frac{\gamma}{2} + n\varepsilon \lg \frac{p}{1-p} = n \left(R + H(p) + \varepsilon \lg \frac{p}{1-p} + \frac{1}{n} \lg \frac{\gamma}{2} \right).$$

Platí $R + H(p) > 1$. Hodnoty $\lg(p/(1-p))$ a $\lg(\gamma/2)$ jsou záporné, ale konstantní. Vhodnou volbou ε a n lze proto učinit odpovídající členy natolik malé, aby celý výraz měl hodnotu přesahující n. □

Zbývá uvést úplný důkaz Shannonovy věty.

Důkaz Věty 1.1. Budeme pracovat s hodnotou $\rho = n(p + \eta)$, kde $\eta > 0$ nebude záviset na n, ale pouze na $\varepsilon > 0$. Jeho volbu upřesníme až v závěrečné části důkazu..

Hledáme binární kód C délky n, který má k kódových slov tak, aby bylo $(\lg k)/n < 1 - H(p)$ a aby metoda volby nejbližšího kódového slova chybovala s libovolně malou předem určenou pravděpodobností $\delta > 0$.

Vyslané kódové slovo budeme opět značit u, zatímco v bude označovat slovo přijaté. Prostor jevů, k němuž se vztahují námi zkoumané a odhadované pravděpodobnosti, je tedy prostor všech možných dvojic (u, v). Případ, kdy se v octne mimo v Hammingově vzdálenosti větší než ρ, budeme považovat bez bližšího zkoumání za případ neúspěšného dekódování. Pravděpodobnost nesprávného dekódování lze tedy shora odhadnout jako

$$\Pr\left(d(u, v) > \rho\right) + \Pr\left((C \setminus \{u\}) \cap S(v, \rho) \neq \emptyset\right).$$

Druhou z uvedených pravděpodobností převedeme na zkoumání, zda dané $u' \in C$ se nachází v blízkosti nějakého přijatého slova v a je přitom různé od u. Probíhá-li u' celý kód C, můžeme součet všech takových pravděpodobností s jistou zkratkovitostí zapsat jako

$$\sum_{u' \neq u} \Pr\left(d(u', v) \leq \rho\right).$$

Touto sumou můžeme nahradit pravý sčítanec ve výše uvedeném odhadu. Při nahrazování jde o horní odhad, neboť v některé kouli se mohou vyskytovat dvě různá u'.

Již jsme dokázali, že ze Zákona velkých čísel 1.2, plyne, že $v \in M(u, \eta)$ pro dostatečně velká n s libovolně velkou pravděpodobností. Pro každé uvažované η lze tedy najít n_0 tak, že pro $n > n_0$ je $\Pr\left(d(u, v) > \rho\right) < \delta/2$.

Abychom vhodně vyjádřili $\sum_{u' \neq u} \Pr\left(d(u', v) \leq \rho\right)$, rozšíříme jevový prostor na trojice (u, u', v), kde $u, u' \in C$, u je vyslané slovo a v je přijaté slovo. Příznivým jevem je ten, kdy $u' \neq u$ a $d(u', v) \leq \rho$. Hledaná pravděpodobnost je pak rovna k-násobku pravděpodobnosti příznivého jevu (připomeňme, že k udává velikost kódu C).

Uvažme chybový vektor $e = v - u$ a položme $w = u' - u$. Máme $d(u', v) \leq \rho$ právě když $d(w, e) \leq \rho$. Místo jevového prostoru daného trojicemi (u, u', v) můžeme tedy vyšetřovat jevový prostor trojic (u, w, e), kde $u \in C$, $w \in u + C$, a příznivým jevem je $w \neq 0$ a $d(w, e) \leq \rho$. Váha jevového podprostoru s pevným e je dána pravděpodobností chyby přesně v nenulových pozicích e, čili je rovna $p^{|e|}(1-p)^{n-|e|}$. Počet příznivých jevů je roven hodnotě

$$\mu_C(e, \rho) = \left|\left\{(u, u') \in C \times C;\ u \neq u' \text{ a } |(u' - u) - e| \leq \rho\right\}\right|,$$

takže pravděpodobnost příznivého jevu pro dané e je rovna $\mu_C(e, \rho)/k^2$. Pro důkaz potřebujeme vyjádření k-násobku souhrnné pravděpodobnosti příznivého jevu, a tou, jak vidíme, je

$$\frac{1}{k} \sum_{e \in \mathbb{F}_2^n} p^{|e|}(1-p)^{n-|e|} \mu_C(e, \rho).$$

Pro dokončení důkazu potřebujeme ověřit, že pro každé dostečně velké n existuje binární kód C s dostatečně velkou nosností $\alpha(C)$, pro který je uvedená hodnota menší než $\delta/2$. Vtip důkazu spočívá v tom, že se kód C nekonstruuje explicitně, ale celá hodnota se počítá tak, jako by $C = \{u_1, \ldots, u_k\}$ bylo proměnným parametrem. Pokud průměr přes všechna C při vhodně zvoleném k vyjde $< \delta/2$, musí samozřejmě nějaké hledané C existovat (a je jich dokonce naprostá většina).

Pro účely výpočtu je výhodné kód C neuvažovat jako k-bodovou množinu, ale jako uspořádanou posloupnost k různých prvků. Všech kódů délky k je $2^n!/(2^n - k)!$. Pokud pro nějaké pevné vektory $u \neq u'$ délky n, požadujeme, aby platilo $u_i = u$ a $u_j = u'$, kde $1 \leq i < j \leq k$, tak je počet takových kódů roven zlomku, kde v čitateli je počet všech kódů a ve jmenovateli je hodnota $2^n(2^n - 1)$. Stejný počet je kódů, kde $u_j = u$ a $u_i = u'$, takže pokud chceme, aby platilo $\{u_i, u_j\} = \{u, u'\}$, musíme do jmenovatele dosadit $\binom{2^n}{2}$. Chceme-li pouze, aby $u, u' \in C$, tak je čitatel třeba vynásobit počtem dvojic $\{i, j\}$, tedy hodnotou $\binom{k}{2}$.

Každá dvojice $\{u, u'\} \subseteq \mathbb{F}_2^n$ se proto vyskytuje právě σ kódech C, kde

$$\sigma = \frac{2^n!\, k(k-1)}{(2^n - k)!\, 2^n(2^n - 1)}.$$

Hodnotu $\sum_C \mu_C(e, \rho)$ můžeme tedy počítat přes všechny dvojice (u, u') z $\mathbb{F}_2^n$, které splňují $u' - u \in S(e, \rho)$ a $u \neq u'$. Takových dvojic je přesně $2^n(V(k, [\rho]) - 1) \leq (2^n - 1)V(k, [\rho]) \leq (2^n - 1)2^{nH(\rho/n)}$, takže $\sum_C \mu_C(e, \rho) \leq \sigma(2^n - 1)2^{nH(\rho/n)}$. Podstatné je, že na pravé straně nefiguruje vektor e. Průměrná hodnota výrazu $\frac{1}{k} \sum_{e \in \mathbb{F}_2^n} p^{|e|}(1-p)^{n-|e|} \mu_C(e, \rho)$ počítaná přes všechny kódy je

tedy shora omezena hodnotou

$$\frac{1}{k}\frac{k(k-1)}{2^n(2^n-1)}(2^n-1)2^{nH(\rho/n)}\sum_{e\in\mathbb{F}_2^n}p^{|e|}(1-p)^{n-|e|}.$$

Závěrečná suma je sumou pravděpodobností přes všechny případy, kterých může nabývat chybový vektor e, a proto je rovna jedné. Získaný odhad lze tedy zjednodušit na

$$\frac{k-1}{2^n}2^{nH(\rho/n)} < k2^{n(H(p+\eta)-1}.$$

Pro dané k a ρ existuje tedy alespoň jeden kód C o k kódových slovech, pro který

$$\sum_{u'\neq u}\Pr\left(d(u',v)\leq\rho\right) < k2^{n(H(p+\eta)-1}.$$

Abychom důkaz dokončili, vyložíme nyní nejprve, že stačí ověřit, že pro každé $\varepsilon>0$ existují $\alpha<0$ a $\eta>0$ taková, že pro každé dostatečně velké n lze najít celé k tak, že pro $R=(\lg k)/n$ platí

$$1-H(p)-\varepsilon\leq R<1-H(p)\ \text{ a }\ R-1+H(p+\eta)\leq\alpha.$$

Podmínka vlevo je jen zopakováním podmínky na nosnost kódu ze znění Věty 1.1. Podmínka vpravo zaručuje, že

$$k2^{n(H(p+\eta)-1}=2^{Rn}2^{n(H(p+\eta)-1}\leq 2^{\alpha n}$$

se s rostoucím n blíží nule a je tedy od některého n menší než $\delta/2$.

Bez újmy na obecnosti můžeme předpokládat, že $p+\eta<1/2$ a $\varepsilon<1-H(p)$. Hodnotu η volíme natolik malou, aby bylo $H(p)+\varepsilon/3\geq H(p+\eta)$. Dále budeme předpokládat, že je $n>3/\varepsilon$. Potom je $n(\varepsilon/3)>1$, takže existuje celé h pro které platí

$$n(1-H(p)-\varepsilon)\leq h\leq n(1-H(p)-2\varepsilon/3).$$

Položme $k=2^h$. Vidíme, že $h=nR$, takže podmínka vlevo splněna je. Splnění druhé podmínky plyne z $h/n\leq 1-H(p)-2\varepsilon/3$, neboť dostáváme

$$R-1+H(p+\eta)\leq H(p+\eta)-H(p)-2\varepsilon/3\leq\varepsilon/3-2\varepsilon/3=-\varepsilon/3.$$

□