SHANNONOVY VĚTY A JEJICH DŮKAZ

JAN ŠŤOVÍČEK

Abstrakt. Podrobnější důkaz Shannonových vět pro binární symetrický kanál. Číslování vět odpovídá přednášce a navazuje na ni.

1. Značení a obecné předpoklady

Pro účely Shannonových vět uvažujeme, že přenos dat probíhá podle následujícího schématu:

$$\boxed{\text{zdroj}} \xrightarrow{X=(X_1,X_2,X_3,\dots)} \boxed{\text{kanál}} \xrightarrow{Y=(Y_1,Y_2,Y_3,\dots)} \boxed{\text{příjemce}}$$

$X_1, X_2, X_3, \dots$ jsou nezávislé náhodné veličiny nabývající hodnot 0 nebo 1 s pravděpodobností $\frac{1}{2}$. Kanál v následujícím uvažujeme pouze binární symetrický s chybovostí p. To jest pravděpodobnost správného přenosu jak nuly tak jedničky je $1-p$:

$$P[Y=0 \mid X=0] = P[Y=1 \mid X=1] = 1-p,$$
$$P[Y=1 \mid X=0] = P[Y=0 \mid X=1] = p.$$

Navíc se jedná o kanál bezpaměťový, tj. pro každé přirozené N a každé dvě posloupnosti bitů $(x_1, x_2, \dots, x_N)$ a $(y_1, y_2, \dots y_N)$ platí

$$P[(Y_1, Y_2, \dots, Y_N) = (y_1, y_2, \dots, y_N) \mid (X_1, X_2, \dots, X_N) = (x_1, x_2, \dots, x_N)] =$$
$$P[Y_1 = y_1 \mid X_1 = x_1] \cdot P[Y_2 = y_2 \mid X_2 = x_2] \cdots P[Y_N = y_N \mid X_N = x_N].$$

Výše uvedené je matematická formulace toho, že výsledek přenosu i-tého bitu nezávisí na výsledku přenosu ostatních bitů.

Pro další diskuzi se vyplatí zavést posloupnost náhodných veličin

$$E = (E_1, E_2, E_3, \dots),$$

kde podle definice $E_i = Y_i - X_i$ (počítáno v tělese $\mathbb{F}_2$). Tato posloupnost náhodných veličin tedy představuje šum kanálu. Z předpokladů na X a Y jednoduše plynou následující fakta:

(1) Náhodné veličiny $X_1, X_2, \dots$ a $E_1, E_2, \dots$ jsou nezávislé.
(2) Pravděpodobnost $P[E_i = 1]$ je rovna p pro všechna i.

To je ostatně přesně ten důvod, proč jsme předpoklady nastavili takto – (1) a (2) nám říkají, že chyby v jednotlivých bitech nastávají s pravděpodobností p, a to nezávisle na chybách v ostatních bitech i na informačním zdroji.

Pro takovýto informační zdroj a kanál jsme spočítali entropie:

$$H(X) = 1 \quad \text{a} \quad H(X \mid Y) = H(p),$$

a tedy vzájemná entropie, kterou lze interpretovat jako informační obsah přijatého symbolu, vyšla

$$I(X,Y) = H(X) - H(X \mid Y) = 1 - H(p).$$

Cílem bude dokázat, že pomocí dostatečně dlouhých kódů s hustotou menší ale libovolně blízkou $1 - H(p)$ můžeme dosáhnout libovolně spolehlivého přenosu (Věta 46), a naopak pro kódy s hustotou větší než $1 - H(p)$ spolehlivost přenosu dlouhých slov klesá k nule (Věta 49).

2. Shannonova věta

Pro formulaci Shannonovy věty potřebujeme připomenout spolehlivost dekódování a spolehlivost kódu.

Definice. Ať $C \subseteq \mathbb{F}_2^N$ je binární kód délky N. Pak *dekódováním* kódu C rozumíme libovolné zobrazení

$$D\colon \mathbb{F}_2^N \longrightarrow C.$$

Dekódování na nejbližší slovo je pak takové, že pro libovolné $v \in \mathbb{F}_2^N$ je Hammingova vzdálenost v a kódového slova $D(v)$ nejmenší možná.

Spolehlivost dekódování D je průměrná pravděpodobnost přes všechna $w \in C$, že při odeslání slova w přes kanál pro přijaté slovo $v \in \mathbb{F}_2^N$ platí $D(v) = w$. Jinak řečeno, jedná se o průměrnou pravděpodobnost, že odeslané slovo bude pomocí D správně dekódováno. V námi zavedeném značení se dá spolehlivost dekódování vyjádřit vzorcem

$$\frac{1}{|C|} \sum_{w \in C} P\big[D\big((Y_1, Y_2, \dots, Y_N)\big) = w \mid (X_1, X_2, \dots X_N) = w\big]$$

Pokud přenášené kódové slovo volíme též náhodně s rovnoměrným rozdělením (tj. pravděpodobnost volby konkrétního $w \in C$ je rovna $\frac{1}{|C|}$), je spolehlivost rovna přesně

$$P\big[D\big((Y_1, Y_2, \dots, Y_N)\big) = (X_1, X_2, \dots X_N)\big],$$

čili pravděpodobnosti toho, že přijaté slovo správně dekódujeme na odeslané.

Nakonec *spolehlivost kódu* C je definována jako maximum spolehlivosti dekódování přes všechna dekódování $D\colon \mathbb{F}_2^N \to C$.

Následující věta nám říká, že pro binární symetrický kanál s chybovostí p existují libovolně spolehlivé kódy s hustotou přibližně $1 - H(p)$. Libovolně spolehlivé je dokonce dekódování na nejbližší slovo, v principu tedy nemusíme vymýšlet žádné komplikovanější dekódovací strategie. Problém z praktického hlediska je ovšem ten, že tyto dobré kódy jsou náhodné kódy. Hledání použitelných kódů dosahujících vlastností předpovězených Shannonovou větou trvalo dalšího půl století a jejich probrání se do této přednášky už nevejde.

Věta 46 (Shannonova věta). *Předpokládejme, že je dán informační zdroj $X = (X_1, X_2, X_3, \dots)$ jako výše a dále binární symetrický kanál s chybovostí p, kde $0 < p < \frac{1}{2}$. Pak pro každá reálná čísla $\varepsilon > 0$ a $\delta > 0$ existuje přirozené číslo N_0 s touto vlastností: Pro každé $N \geq N_0$ existuje binární kód C délky N s hustotou alespoň $1 - H(p) - \varepsilon$, pro který je spolehlivost dekódování na nejbližší slovo alespoň $1 - \delta$.*

Intuitivně víme, že při odeslání N bitů a pravděpodobnosti chyby p by mělo nastat přibližně $p \cdot N$ chyb. Tuto skutečnost formálně vyjadřuje tzv. slabý zákon velkých čísel, který uvedeme bez důkazu.

Věta 47 (Slabý zákon velkých čísel). *Ať $Z_1, Z_2, Z_3, \dots$ je posloupnost náhodných veličin se stejným rozdělením taková, že každé Z_i nabývá jednu z konečně mnoha reálných hodnot $a_1, a_2, \dots, a_m$, a to s pravděpodobnostmi $p_1, p_2, \dots, p_m$. Položme $\mu = \sum_{j=1}^m p_j a_j$ (tj. μ je střední hodnota Z_i pro libovolné i). Pak pro libovolné $\varepsilon > 0$ platí*

$$\lim_{N\to\infty} P\left[\left|\frac{1}{N}\sum_{i=1}^N Z_i - \mu\right| \geq \varepsilon\right] = 0$$

.

V našem případě, dosadíme-li za Z_i náhodnou veličinu šumu E_i, je $\sum_{i=1}^N E_i$ rovno počtu chyb při přenosu N bitů. Je-li tedy w vyslané a v přijaté slovo délky N, platí, že

$$P\big[|d(w,v) - pN| \geq \varepsilon N\big]$$

jde k nule pro $N \to \infty$.

Poznámka. Jelikož jsou $E_1, E_2, E_3, \dots$ podle předpokladu nezávislé náhodné veličiny, platí dokonce silnější Černovova nerovnost, která říká, že

$$P\big[d(w,v) \geq (p+\varepsilon)N\big] \leq e^{-N\varepsilon^2/2}.$$

Nyní můžeme Shannonovu větu dokázat.

Důkaz Věty 74. Řekněme, že máme dány $\varepsilon > 0$ a $\delta > 0$. Jak již bylo řečeno, kandidáty na kód $C \subseteq \mathbb{F}_2^N$, kde N je dostatečně velké a dekódování na nejbližší slovo má spolehlivost alespoň $1 - \delta$, volíme náhodně. Minimální délku N_0 takovýchto kódů upřesníme dále v průběhu důkazu. Počet slov M kódu C délky N budeme volit tak, aby platilo

$$1 - H(p) - \varepsilon < \frac{\log_2 M}{N} < 1 - H(p) - \frac{\varepsilon}{2}. \tag{$*$}$$

První nerovnost je vyžadována ve znění věty, druhou potřebujeme pro důkaz. Pro dostatečně velké N nějaké přirozené číslo M s těmito vlastnostmi jistě najdeme.

Nyní si uvedeme pár podrobností k tomu, jak pro danou délku N a počet slov M kód C volíme. Z technických důvodů budeme C považovat za náhodnou veličinu, která s rovnoměrným rozdělením nabývá všech *uspořádaných* M-tic

$$(c_1, c_2, \dots, c_M)$$

po dvou různých slov délky N v abecedě $\mathbb{F}_2$. Dále požadujeme, aby náhodná veličina C byla nezávislá na šumu kanálu E.

Jednotlivé složky veličiny C označíme $C_1, C_2, \dots, C_M$. To jest, C_i je i-té slovo našeho náhodně zvoleného kódu C. Jednoduché cvičení na počítání s pravděpodobnostmi nám pro každou dvojici různých čísel $i, j \in \{1, 2, \dots, M\}$ a dvojici libovolných slov $c, d \in \mathbb{F}_2^N$ dá následující:

- $P[C_i = c] = 2^{-N}$. Jinak řečeno, pro libovolné i nabývá C_i slov délky N s rovnoměrným rozdělením.

- $P[C_i = c \;\&\; C_j = d] = \frac{1}{2^N(2^N-1)}$. Mimo jiné odtud okamžitě plyne něco, na co musíme dát pozor: Pro různá i a j *nejsou* C_i a C_j nezávislé náhodné veličiny.

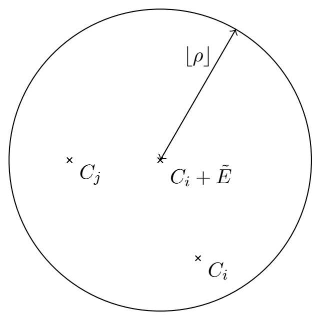
Přistoupíme k dalšímu kroku důkazu a odhadneme, jaká je pro danou délku N spolehlivost dekódování na nejbližší slovo pro *náhodný* kód C zvolený výše uvedeným způsobem. Řekněme, že jsme odeslali i-té kódové slovo C_i (které je nyní samo o sobě náhodnou veličinou!) a přijali slovo $C_i + \tilde{E}$ zatížené náhodnou chybou $\tilde{E} = (E_1, E_2, \ldots, E_N)$. Případy, kdy došlo k chybnému dekódování, si rozdělíme do dvou skupin. Pro toto dělení použijeme malé kladné reálné číslo $\eta > 0$, jehož hodnotu upřesníme později. Případy chybného dekódování se rozpadnou do následujících skupin:

a) Došlo k příliš mnoho chybám. V našem případě zvolíme pro “příliš mnoho” mez $\rho = (p + \eta)N$ a pokud došlo při přenosu k více než ρ chybám, nebudeme důvod neúspěšného dekódování dále rozebírat. To proto, že ze slabého zákona velkých čísel platí pro dostatečně velké N:

$$P\big[w(\tilde{E}) > \rho\big] < \frac{\delta}{2}$$

Všechny případy, kdy došlo k více než ρ chybám, ať už dekódování dopadlo jakkoli, mají tedy souhrnnou pravděpodobnost menší než $\frac{\delta}{2}$.

b) Došlo k nejvýše ρ chybám, ale přesto jsme přijaté slovo chybně dekódovali. Jak je vidět z následujícího obrázku, to se mohlo stát pouze v případě, že kombinatorická koule $B(C_i + \tilde{E}, \lfloor \rho \rfloor)$ obsahuje nějaké další kódové slovo, tj. C_j pro nějaké $j \neq i$.



Obrázek: Chybné dekódování pro $w(\tilde{E}) \leq \rho$.

Pravděpodobnost chybného dekódování v tomto případě je tedy omezena hodnotou

$$P\big[(\exists j \in \{1, \ldots, i-1, i+1, \ldots, M\})(d(C_i + \tilde{E}, C_j) \leq \rho)\big].$$

Tato pravděpodobnost je pak zjevně shora omezena součtem všech pravděpodobností, že k situaci podobné té na obrázku dojde pro každé jedno dané $j \in \{1, \ldots, i-1, i+1, \ldots, M\}$, tj. součtem

$$\sum_{\substack{j=1,\ldots,M \\ j \neq i}} P\big[d(C_i + \tilde{E}, C_j) \leq \rho\big].$$

Z diskuze případů a) a b) vyplývá, že pro to, abychom pravděpodobnost chybného dekódování odhadli shora konstantou δ, stačí, abychom v případě

b) dokázali (pro dostatečně velkou délku kódu N a vhodnou volbu čísla $\eta > 0$) odhad

$$\sum_{\substack{j=1,\dots,M \\ j\neq i}} P\big[d(C_i+\tilde{E}, C_j) \leq \rho\big] < \frac{\delta}{2}. \tag{†}$$

Nezbude, než provést provést několik výpočtů. Předně můžeme pravděpodobnost $P\big[d(C_i+\tilde{E}, C_j) \leq \rho\big]$ rozdělit podle hodnoty, kterou nabude $\tilde{E}$:

$$P\big[d(C_i+\tilde{E}, C_j) \leq \rho\big] = \sum_{e\in\mathbb{F}_2^N} P\big[d(C_i+e, C_j) \leq \rho \mid \tilde{E}=e\big] \cdot P\big[\tilde{E}=e\big]$$

Ze společného rozdělení C_i a C_j a jejich nezávislosti na $\tilde{E}$ pak plyne, že

$$\sum_{e\in\mathbb{F}_2^N} P\big[d(C_i+e, C_j) \leq \rho \mid \tilde{E}=e\big] \cdot P\big[\tilde{E}=e\big] =$$
$$\sum_{e\in\mathbb{F}_2^N} P\big[d(C_i+e, C_j) \leq \rho\big] \cdot P\big[\tilde{E}=e\big] =$$
$$\sum_{e\in\mathbb{F}_2^N} \frac{\big|\{(c,d)\in\mathbb{F}_2^N\times\mathbb{F}_2^N \mid d(c+e,d)\leq\rho\}\big|}{2^N(2^N-1)} \cdot P\big[\tilde{E}=e\big]$$

Velikost množiny $\{(c,d)\in\mathbb{F}_2^N\times\mathbb{F}_2^N \mid d(c+e,d)\leq\rho\}$ vypočítáme následovně. Slovo c si můžeme zvolit libovolně z 2^N možností. Slovo d pak musí být různé od c a musí padnout do kombinatorické koule $B(c+e, \lfloor\rho\rfloor)$, máme tedy $V(N,\lfloor\rho\rfloor)-1$ různých možností. Zkoumaná množina má proto $2^N\cdot\big(V(N,\lfloor\rho\rfloor)-1\big)$ prvků a máme:

$$P\big[d(C_i+\tilde{E}, C_j) \leq \rho\big] =$$
$$\sum_{e\in\mathbb{F}_2^N} \frac{\big|\{(c,d)\in\mathbb{F}_2^N\times\mathbb{F}_2^N \mid d(c+e,d)\leq\rho\}\big|}{2^N(2^N-1)} \cdot P\big[\tilde{E}=e\big] =$$
$$\sum_{e\in\mathbb{F}_2^N} \frac{2^N\big(V(N,\lfloor\rho\rfloor)-1\big)}{2^N(2^N-1)} \cdot P\big[\tilde{E}=e\big] = \frac{2^N\big(V(N,\lfloor\rho\rfloor)-1\big)}{2^N(2^N-1)} \cdot \sum_{e\in\mathbb{F}_2^N} P\big[\tilde{E}=e\big] =$$
$$\frac{2^N\big(V(N,\lfloor\rho\rfloor)-1\big)}{2^N(2^N-1)}$$

Jednoduše nahlédneme, že $2^N\cdot\big(V(N,\lfloor\rho\rfloor)-1\big) \leq (2^N-1)\cdot V(N,\lfloor\rho\rfloor)$, a tedy platí

$$P\big[d(C_i+\tilde{E}, C_j) \leq \rho\big] \leq \frac{(2^N-1)V(N,\lfloor\rho\rfloor)}{2^N(2^N-1)} = \frac{V(N,\lfloor\rho\rfloor)}{2^N}.$$

Pokud jsme zvolili konstantu η dostatečně malou tak, aby $p+\eta<\frac{1}{2}$, pak $\rho=(p+\eta)N<\frac{N}{2}$. Můžeme tedy velikost $V(N,\lfloor\rho\rfloor)$ kombinatorické koule o poloměru $\lfloor\rho\rfloor$ odhadnout pomocí lemmatu 36 dříve z přednášky:

$$V(N,\lfloor\rho\rfloor) \leq 2^{N\cdot H(\frac{\lfloor\rho\rfloor}{N})} \leq 2^{N\cdot H(\frac{\rho}{N})} = 2^{N\cdot H(p+\eta)}.$$

Druhá nerovnost opět plyne z toho, že $\rho/N < \frac{1}{2}$ a entropická funkce je na intervalu $(0, \frac{1}{2})$ rostoucí. Dohromady pak máme

$$P\big[d(C_i + \tilde{E}, C_j) \le \rho\big] \le \frac{2^{N \cdot H(p+\eta)}}{2^N} = 2^{N(H(p+\eta)-1)}.$$

Nyní se vrátíme k důkazu nerovnosti (†) na stránce 5. Právě jsme dokázali odhad pro každý člen sumy na levé straně, čili platí:

$$\sum_{\substack{j=1,\dots,M \\ j \ne i}} P\big[d(C_i + \tilde{E}, C_j) \le \rho\big] \le (M-1) \cdot 2^{N(H(p+\eta)-1)} < M \cdot 2^{N(H(p+\eta)-1)}.$$

Počet slov jsme volili tak, aby platily nerovnosti (∗) na straně 3, speciálně tedy aby

$$M < 2^{N(1-H(p)-\frac{\varepsilon}{2})}.$$

Po dosazení této nerovnosti do předchozí dostaneme

$$\sum_{\substack{j=1,\dots,M \\ j \ne i}} P\big[d(C_i + \tilde{E}, C_j) \le \rho\big] \le 2^{N(H(p+\eta)-H(p)-\frac{\varepsilon}{2})}.$$

Jelikož H je v bodě p spojitá, což se dá vyjádřit tím, že $\lim_{\eta \to 0} H(p+\eta) = H(p)$, bude pro dostatečně malé kladné číslo η platit

$$H(p+\eta) - H(p) - \frac{\varepsilon}{2} < 0.$$

Při volbě vhodného η je tedy

$$\sum_{\substack{j=1,\dots,M \\ j \ne i}} P\big[d(C_i + \tilde{E}, C_j) \le \rho\big] \le 2^{N\gamma}$$

pro nějaké $\gamma < 0$ a pravá strana se s rostoucím N blíží k nule. Speciálně platí pro N dostatečně velké, že

$$\sum_{\substack{j=1,\dots,M \\ j \ne i}} P\big[d(C_i + \tilde{E}, C_j) \le \rho\big] < \frac{\delta}{2}$$

což jsme měli dokázati.

Shrňme si tedy dosavadní postup. Dokázali jsme, že zvolíme-li dostatečně dlouhý kód C o M slovech s hustotou mezi $1 - H(p) - \varepsilon$ a $1 - H(p) - \frac{\varepsilon}{2}$ náhodně, pak pravděpodobnost správného dekódování libovolného zvoleného slova C_i je větší než $1-\delta$. Speciálně tedy průměrná pravděpodobnost správného dekódování braná přes všech M kódových slov (tedy vlastně spolehlivost dekódování na nejbližší slovo pro náhodný kód) je větší než $1-\delta$.

Protože jsme C volili náhodně s rovnoměrným rozdělením, je tato průměrná pravděpodobnost ve skutečnosti průměrem spolehlivostí všech $\frac{(2^N)!}{(2^N-M)!}$ posloupností $(c_1, c_2, \dots, c_M)$, kterých může náhodná veličina C nabývat. Jinak řečeno, průměr spolehlivosti dekódování na nejbližší slovo přes *všechny* možné kódy o M slovech je větší než $1-\delta$. Musí tedy existovat alespoň jeden konkrétní kód, který má požadovanou spolehlivost dekódování na nejbližší slovo větší než $1-\delta$, což jsme měli dokázat. □

3. Obrácená Shannonova věta

Shannonova věta nám říká, že pro přenos přes binární symetrický kanál s chybovostí p existují dobré kódy s hustotou menší než informační obsah přijatého bitu $I(X,Y) = 1 - H(p)$, ale libovolně blízkou tomuto obsahu. Obrácená Shannonova věta nám říká, že hustotu $1 - H(p)$ nemůžeme při volbě dobrých kódů překročit. Nejdříve si dokážeme jednoduché lemma:

Lemma 48. *Jsou-li A, B dva jevy takové, že A nastává s pravděpodobností a a B nastává s pravděpodobností b, pak $A \cap B$ nastává s pravděpodobností alespoň $a + b - 1$.*

Důkaz. Označme $x = P[A \cap B]$. Pak

$$a - x = P[A \setminus B] \leq P[\text{nenastane } B] = 1 - b.$$

Odtud $x \geq a + b - 1$. □

Nyní formulujeme a dokážeme slíbenou větu. Na rozdíl od Věty 74, kde stačilo uvažovat dekódování na nejbližší slovo, nás v následujícím případě od špatné spolehlivosti nezachrání ani sebepropracovanější metoda dekódování.

Věta 49 (Obrácená Shannonova věta). *Ať $\delta > 0$ a $p \in (0, \frac{1}{2})$ jsou reálná čísla a uvažujme binární symetrický kanál s chybovostí p. Pak pro každé $R > 1 - H(p)$ existuje přirozené číslo N_0 takové, že každý binární kód délky $N \geq N_0$ s hustotou alespoň R má spolehlivost nejvýše δ.*

Důkaz. Pro dané p pro spor předpokládejme opak. Tedy že existují $\delta > 0$ a $R > 1 - H(p)$, pro které není délka kódů C s hustotou alespoň R a spolehlivostí větší než δ omezena žádným N_0. Uvažujme jeden takový kód C, jehož délku N upřesníme později, spolu s nějakým dekódováním $D\colon \mathbb{F}_2^N \to C$ spolehlivosti alespoň δ.

Ať $\varepsilon > 0$ je kladná reálná konstanta, jejíž hodnotu upřesníme dále. Zatím požadujeme pouze, aby $\varepsilon < p$. Zvolme náhodně s rovnoměrným rozdělením a nezávisle na šumu kanálu E slovo $w \in C$. Pošleme-li toto slovo po kanále, přijmeme slovo $v = w + \tilde{E}$, kde $\tilde{E} = (E_1, E_2, \ldots, E_N)$. Podle lemmatu 76 můžeme odhadnout

$$P\big[v \in D^{-1}(w) \;\&\; |d(w,v) - pN| < \varepsilon N\big] \geq \\ P\big[v \in D^{-1}(w)\big] + P\big[|d(w,v) - pN| < \varepsilon N\big] - 1.$$

První pravděpodobnost v součtu je pravděpodobnost, že slovo v bude pomocí D správně dekódováno zpět na w. Jedná se tedy o spolehlivost dekódování D a podle předpokladu je proto $P\big[v \in D^{-1}(w)\big] > \delta$. Podle věty 75 je pro dostatečně velké N druhá pravděpodobnost v součtu větší než $1 - \delta/2$. Dohromady proto

$$P\big[v \in D^{-1}(w) \;\&\; |d(w,v) - pN| < \varepsilon N\big] > \delta + (1 - \frac{\delta}{2}) - 1 = \frac{\delta}{2}. \qquad (\ddagger)$$

Nyní odhadneme poslední pravděpodobnost jinak. Zvolme pevné $w \in C$ a počítejme pravděpodobnost jevu $\{v \in D^{-1}(w) \;\&\; |d(w,v) - pN| < \varepsilon N\}$ za předpokladu, že bylo odesláno toto konkrétní w. Dostaneme:

$$P\big[v \in D^{-1}(w) \;\&\; |d(w,v) - pN| < \varepsilon N \mid \text{odesláno } w\big] = \sum_{v \in M_w} P[\tilde{E} = w - v],$$

kde množina M_w je definována jako

$$M_w = \{v \in \mathbb{F}_2^N \mid v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N\}.$$

Tvrdíme, že pro $v \in M_w$ platí $P[\tilde{E} = w - v] \leq 2^{-NH(p)} \left(\frac{1-p}{p}\right)^{\varepsilon N}$. Všimněme si, že pokud toto tvrzení dokážeme, pak automaticky

$$P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N \mid \text{odesláno } w\big] \leq |M_w| \cdot 2^{-NH(p)} \left(\frac{1-p}{p}\right)^{\varepsilon N}.$$

Tvrzení nahlédneme následovně. Slovo $w - v$ je v $P[\tilde{E} = w - v]$ pevně zvolené, a protože chyby v jednotlivých bitech nastávají nezávisle na sobě s pravděpodobností p, platí

$$P[\tilde{E} = w - v] = p^{d(w,v)}(1-p)^{N-d(w,v)}.$$

Jelikož máme pro $v \in M_w$ nerovnost $d(w,v) \geq (p-\varepsilon)N$ a funkce $p^t(1-p)^{N-t}$ je pro $t \in (0,N)$ klesající, je

$$P[\tilde{E} = w - v] \leq p^{(p-\varepsilon)N}(1-p)^{N-(p-\varepsilon)N} = p^{pN}(1-p)^{(1-p)N}\left(\frac{1-p}{p}\right)^{\varepsilon N}.$$

Z definice $H(p)$ přímo plyne, že $p^{pN}(1-p)^{(1-p)N} = 2^{-NH(p)}$, čímž je tvrzení dokázáno.

Volíme-li nyní $w \in C$ opět náhodně, máme s výše vypočítaným odhad:

$$\begin{aligned}
&P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N\big] = \\
&\sum_{w \in C} P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N \mid \text{odesláno } w\big] \cdot P\big[\text{odesláno } w\big] = \\
&\frac{1}{|C|} \cdot \sum_{w \in C} P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N \mid \text{odesláno } w\big] \leq \\
&\frac{1}{|C|} \cdot 2^{-NH(p)} \left(\frac{1-p}{p}\right)^{\varepsilon N} \cdot \sum_{w \in C} |M_w|.
\end{aligned}$$

Protože jsou pro různá $w \in C$ množiny $D^{-1}(w)$, a tedy i množiny M_w, disjunktní, platí $\sum_{w \in C} |M_w| \leq 2^N$. Dostaneme proto

$$P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N\big] \leq \frac{1}{|C|} \cdot 2^{N(1-H(p))} \left(\frac{1-p}{p}\right)^{\varepsilon N}.$$

Podle předpokladu je hustota kódu C alespoň R, tj. $|C| \geq 2^{RN}$ a máme

$$P\big[v \in D^{-1}(w) \ \&\ |d(w,v) - pN| < \varepsilon N\big] \leq 2^{N(1-H(p)-R)} \left(\frac{1-p}{p}\right)^{\varepsilon N}.$$

Kombinujeme-li poslední nerovnost s nerovností (‡) na straně 7, dostaneme

$$\frac{\delta}{2} < 2^{N(1-H(p)-R)} \left(\frac{1-p}{p}\right)^{\varepsilon N}.$$

Po zlogaritmování a vydělení délkou N má tato nerovnost tvar:

$$0 < 1 - H(p) - R + \varepsilon \log_2 \frac{1-p}{p} - \frac{1}{N} \log_2 \frac{\delta}{2}.$$

Tato nerovnost má platit pro libovolně velké délky kódů N a pro libovolně malé hodnoty ε. Ukážeme, že to není možné. Z předpokladu totiž $R > 1 - H(p)$, čili $1 - H(p) - R$ je záporné číslo. Vhodnou volbou ε a N můžeme docílit toho, že součet

$$\varepsilon \log_2 \frac{1-p}{p} - \frac{1}{N} \log_2 \frac{\delta}{2}$$

má libovolně malou absolutní hodnotu. Speciálně je tedy pro dostatečně velké N a dostatečně malé ε hodnota výrazu

$$1 - H(p) - R + \varepsilon \log_2 \frac{1-p}{p} - \frac{1}{N} \log_2 \frac{\delta}{2}$$

záporná, což je hledaný spor. $\square$