1. Shannonova věta pro binární symetrický kanál

Věta. Nechť je dán binární symetrický kanál s chybovostí p, kde $0 < p < \frac{1}{2}$. Pro každé $\varepsilon > 0$ lze přes kanál přenášet kódem délky n informaci rychlostí $1 - H(p) - \varepsilon$ bitu na jeden symbol s pravděpodobností chyby, která se s rostoucím n blíží nule.

Důkaz. Pro dosažení ohlášené rychlosti přenosu je třeba přenést jedním kódovým slovem délky n jednu z m stejně pravděpodobných zpráv, kde $m \geq 2^{n(1-H(p)-\varepsilon)}$. Zvolme m tak, aby

$$2^{n(1-H(p)-\frac{\varepsilon}{2})} \geq m \geq 2^{n(1-H(p)-\varepsilon)}$$

(pro velká n je to jistě možné). Přiřaďme dále i-té zprávě náhodně zvolené binární slovo c_i délky n. (Může se tedy stát, že dvě různé zprávy jsou reprezentovány stejným slovem).

Zvolme $\eta > 0$ takové, že $p+\eta < \frac{1}{2}$ a $H(p+\eta)-H(p) < \frac{\varepsilon}{2}$. Položme $r = \lfloor n(p+\eta) \rfloor$.

Přijaté slovo y budeme dekódovat takto: pokud v kouli $B(y,r)$ leží kódové slovo c_i pro jediné i, dekódujeme na i-tou zprávu, jinak vyhlásíme selhání.

Existují dvě možnosti chybného dekódování (které se vzájemně nevylučují):

(1) vyslané slovo neleží v kouli $B(y,r)$, tj. došlo k více než r chybám;
(2) v kouli leží c_i pro nějaké i, které nebylo kódovanou zprávou.

Pravděpodobnost první možnosti je omezena Černovovou nerovností a pro velká n se blíží k nule.

Pravděpodobnost druhé možnosti shora odhadneme jako

$$\sum_{i=1}^{m} \frac{V(n,r)}{2^n} < m \cdot 2^{nH(\frac{r}{n})} \cdot 2^{-n} \leq 2^{n(1-H(p)-\frac{\varepsilon}{2})} \cdot 2^{nH(p+\eta)} \cdot 2^{-n} = \\ = 2^{n(H(p+\eta)-H(p)-\frac{\varepsilon}{2})}.$$

I tato pravděpodobnost se blíží nule, čímž je věta dokázána. □

V důkazu věty jsme nepoužili žádný explicitní kód. Navíc jsme pracovali s multikódem: kódová slova se mohla opakovat. Pokud bychom chtěli ukázat, že existuje alespoň jeden konkrétní kód délky n s m slovy splňující podmínky, můžeme uvažovat takto: Náhodná volba kódových vede celkově k pravděpodobnosti neúspěchu nejvýše δ_n (kde δ_n je mez chyby pro dané n jdoucí pro velká n k nule). Existuje tedy alespoň jeden konkrétní multikód s takto omezenou průměrnou chybovostí. Opakující se slova nyní můžeme nahradit libovolně zvolenými dosud nepoužitými slovy a při dekódování tato slova (říkejme jim *duplicitní*) ignorovat. Je snadno vidět, že takovým postupem se chybovost nezvýší: je-li zvoleno k přenosu jedno z duplicitních slov, dochází k chybě dekódování, stejně jako u multikódu. V ostatních případech se chybovost dokonce mírně sníží, protože přítomnost duplicitních slov v dekódované kouli je ignorována.

2. Inverzní Shannonova věta pro binární symetrický kanál

Věta. Nechť je dán binární symetrický kanál s chybovostí p, kde $0 < p < \frac{1}{2}$. Je-li pro $\varepsilon > 0$ přes kanál přenášena kódem délky n informace rychlostí alespoň $1 - H(p) + \varepsilon$ bitu na jeden symbol, klesá s rostoucím n pravděpodobnost správného dekódování k nule.

Pravdivost věty plyne z následující úvahy: Úspěšné dekódování zprávy vede ke zjištění nastalé chyby, což je nezávislý zdroj informace s entropií $H(p)$. Celkem tedy

při správném dekódování získáme $1+\varepsilon$ bitu na jeden binární symbol. Z toho je vidět, že nejistota při dekódování je alespoň o $n \cdot \varepsilon$ bitech informace, a pravděpodobnost jejich správného určení (uhodnutí) se s rostoucím n blíží nule.

Ukažme nyní rigorózní důkaz.

Důkaz. Úspěšné dekódování sestává ze tří náhodných procesů:

- volba kódového slova $w \in C$;
- volba chybového slova $e \in \{0,1\}^n$;
- algoritmus dekódování.

Označme W náhodnou veličinu nabývající hodnoty vybraného kódového slova. Budeme předpokládat, že W nabývá uniformně náhodně hodnotu jednoho z m binárních kódových slov délky n, kde

$$m \geq 2^{1-H(p)+\varepsilon}. \tag{1}$$

Takovou situaci často předpokládáme tiše. Že tento předpoklad není na újmu obecnosti plyne z toho, že neuniformní volba umožňuje kompresi. Jinak řečeno, neuniformně rozdělené zprávy lze zakódovat uniformně rozdělenými zprávami menší délky (tomu říkáme „kódování zdroje“).

Náhodnou veličinu chyby označme E. Rozdělení E je charakteristika kanálu. Přesněji, kanál je charakterizován souborem pravděpodobností $\Pr[Y = y \mid W = w]$, kde Y je náhodná veličina výstupu kanálu. Jde tedy o pravděpodobnosti, že při vyslání zprávy w je na konci kanálu doručena zpráva y. V případě BSC platí $Y = W + E$, kde $E = e$ je uniformně náhodná veličina nabývající hodnot z $\{0,1\}^n$.

Dekódovací algoritmus musíme nechat neupřesněný, tvrdíme totiž, že dekódování bude málo úspěšné pro všechny možné dekódovací strategie. Nemůžeme dokonce předpokládat ani to, že dekódování je deterministické. Podobně jako v případě kanálu je tedy dekódování charakterizováno souborem podmíněných pravděpodobností $\Pr[D(y) = \hat{y}] := \Pr[\hat{Y} = \hat{y} \mid Y = y]$, tedy pravděpodobnosí, že je-li na vstupu dekódovacího algoritmu y, bude výsledkem dekódování $\hat{y}$. Všimněme si, že $Y = W + E$, ale $\Pr[D(y) = \hat{y}]$ je závislá pouze na y, nikoli na tom, jak y vzniklo jakožto součet $w + e$ (podstata šumu je právě to, že vypadá stejně jako zpráva).

Pravděpodobnost úspěšného dekódování $P = \Pr[W = w, E = e, D(e + w) = w]$ je tedy díky uvedené nezávilosti

$$\begin{aligned} P &= \sum_{w \in C} \sum_{e \in \{0,1\}^n} \Pr[W = w] \cdot \Pr[E = e] \cdot \Pr[D(w + e) = w] = \\ &= \frac{1}{m} \sum_{w \in C} \sum_{e \in \{0,1\}^n} \Pr[E = e] \cdot \Pr[D(w + e) = w]. \end{aligned}$$

Chyby rozdělíme podle váhy na dvě části s pomocí kladné konstanty η splňující

$$\left(\frac{1-p}{p}\right)^{\eta} < 2^{\frac{\varepsilon}{2}}.$$

Nechť E_1 je množina chyb váhy nejvýše $n \cdot (p - \eta)$ a E_2 množina chyb větší váhy. Pak $P = P_1 + P_2$, kde

$$P_i = \frac{1}{m} \sum_{w \in C} \sum_{e \in E_i} \Pr[E = e] \cdot \Pr[D(w + e) = w].$$

Pro P_1 dostáváme

$$P_1 \leq \frac{1}{m} \sum_{w \in C} \sum_{e \in E_1} \Pr[E = e] = \Pr[E \in E_1],$$

což jde podle Černovova odhadu s rostoucím n k nule.

Pravděpodobnost, že chyba je v E_2, lze shora odhadnout chybami s nejmenší vahou, protože $p < \frac{1}{2}$. Tedy

$$\Pr[E \in E_2] < p^{n \cdot (p-\eta)} (1-p)^{n(1-p+\eta)} = 2^{-nH(p)} \cdot \left(\frac{1-p}{p} \right)^{n \cdot \eta} < 2^{n(\frac{\varepsilon}{2} - H(p))}. \tag{2}$$

Všimněme si, že pro libovolné $y \in \{0,1\}^n$ platí

$$\sum_{w \in \{0,1\}^n} \Pr[D(y) = w] \leq 1\,,$$

s rovností, pokud dekódovací algoritmus vždy vrací smysluplný výstup (např. nikdy nehlásí selhání). Můžeme tedy psát

$$\sum_{w \in C} \sum_{e \in E_2} \Pr[D(w+e) = w] \leq \sum_{y \in \{0,1\}^n} \sum_{w \in C} \Pr[D(y) = w] \leq 2^n. \tag{3}$$

(Obě nerovnosti jsou rovnostmi v obvyklém případě, že algoritmus vždy vrací kódové slovo.) Z (1), (2) a (3) nyní dostáváme

$$P_2 < 2^{n(H(p)-1-\varepsilon)} \cdot 2^{n(\frac{\varepsilon}{2} - H(p))} \cdot 2^n = 2^{-n\frac{\varepsilon}{2}}.$$

Tedy i P_2 jde s rostoucím n k nule a důkaz je hotov.

□