

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

29. dubna 2020

- Entropie dává dolní odhad na délku kódu (bez šumu)
- Entropie dává horní odhad na objem koule v $\{0, 1\}^n$
- Binární symetrický kanál o chybovosti p , kapacitě $1 - H(p)$
- Kód, který se vždy dekóduje správně, má hustotu shora omezenou kapacitou kanálu
- Problém: Při chybovosti kanálu $\in (0, 1)$ se každý kód občas dekóduje špatně

- Náhodně vyberu zdrojové slovo
- Máme BSC s chybovostí $p = 1/2$
- Pro každé $n \in \mathbb{N}$ a každou volbu blokového (binárního) kódu C délky n bude pravděpodobnost, že správně dekódujeme nejvýše $1/|C|$.
- Tj. dekódování není lepší než hádání

Komunikace skrz kanál s $p = 1/2$

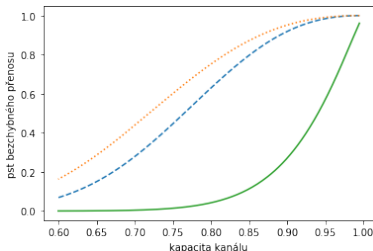
- Posílám $a \in \{0, 1\}^n$, kanál generuje $c \in \{0, 1\}^n$, přijímám $b = a + c$
- Mějme kód C a dekodovací funkci D
- Jaká je pro dané $a \in C$ pst, že $D(a + c) = a$?
- Předpokládejme, že jsme odeslali a , jak vypadá c ?
- Chybovost $p = 1/2$, tedy pro každý prvek $x \in \{0, 1\}^n$ je pst $c = x$ přesně 2^{-n}
- Tedy: pro každý prvek $b \in \{0, 1\}^n$ je pst $b = a + c$ přesně 2^{-n}

- Tedy: pro každý prvek $b \in \{0, 1\}^n$ je pst $b = a + c$ přesně 2^{-n}
- Pravděpodobnostní rozdělení přijatého slova nezávisí na volbě a
- Značme $D^{-1}(a)$ množinu všech b , že $D(b) = a$
- Tedy pro $a \in C$ pst, že $D(a + c) = a$ je $2^{-n}|D^{-1}(a)|$
- Pozorování: $\sum_{a \in C} |D^{-1}(a)| \leq 2^n$
- V průměru je pravděpodobnost správného dekódování

$$\frac{1}{|C|} \sum_{a \in C} 2^{-n}|D^{-1}(a)| \leq \frac{1}{|C|} 2^{-n} 2^n = \frac{1}{|C|}$$

- Kdysi dávno jsme měli BSC s chybovostí $p = 1\%$
- Entropie $H(0,01) \doteq 0,08$, kapacita kanálu asi 0,92
- Totální kód: Hustota 1, pst. chyby asi 63% na 100znakové zdrojové slovo
- Uvážili jsme opakovací kód délky 3 – hustota $1/3 \doteq 0,33$, pst. chyby asi 3% na 100znakové zdrojové slovo
- Dále jsme uvážili kód H_3 – hustota $4/7 \doteq 0,57$, pst. chyby asi 5% na 100znakové zdrojové slovo

Jak závisí pst přenosu 100 znaků bez chyby na kapacitě kanálu

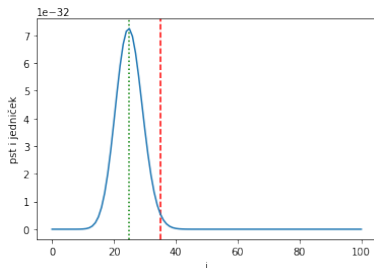


- Zelená plná: Totální kód
- Tečkovaná oranžová: Opakovací kód
- Modrá přerušovaná: H_3

S kolika chybami počítat

Theorem

Bud' $p \in [0, 1], n \in \mathbb{N}, \alpha > 0$. Necht' $z_1, \dots, z_n \in \{0, 1\}$ jsou volené nezávisle tak, že $\text{pst } z_i = 1$ je p . Pak pravděpodobnost, že $\sum_{i=1}^n z_i \geq n(p + \alpha)$ je nejvýše $e^{-n\alpha^2/2}$.



- Obrázek pro $n = 100, p = 1/4, \alpha = 1/10$
- Pro velká n to plyne ze zákona velkých čísel
- Důsledek Čebyševovy nerovnosti (Kaiser Věta 11.2.1)

- Lze pro každou chybovost p najít kód, který se přesně trefí do kapacity kanálu a bude mít spolehlivost skoro 1?
- **Spolehlivost kódu** C při chybovosti kanálu p definujeme jako

$$\rho_C(p) = \frac{1}{|C|} \sum_{\gamma \in C} P[\text{odeslané slovo } \gamma \text{ je správně dekodováno}]$$

- **Pozor:** Na rozdíl od předchozích úvah není spolehlivost normována na zdrojové slovo stejné délky
- Ukážeme, že pro každý kanál existují spolehlivé kódy s hustotou zhruba kapacita kanálu

Theorem

Nechť $p \in (0, 1/2)$. Mějme kanál s chybovostí p a kapacitou $C(p) = 1 - H(p)$. Nechť $\kappa < C(p)$. Potom pro každé $\varepsilon > 0$ existuje binární kód s hustotou k/n aspoň κ a spolehlivostí $> 1 - \varepsilon$.

- Oproti skriptům drobně přeznačeno.
- Měli byste už chápat, proč $p = 0$, $p = 1/2$ a $p > 1/2$ nejsou zajímavé
- Z důkazu bude vidět, že dokonce pro každé $\varepsilon > 0$ existuje n_0 , že pro $n \geq n_0$ existuje kód délky n splňující podmínky
- Trikový důkaz: Není to „ C zkonstruujeme tak a tak“, ale „volme C náhodně a ukážeme, že často má dost dobré vlastnosti“

- Abychom ukázali, že existuje dobrý C , zvolíme C náhodně a ukážeme, že pravděpodobnost jevu „ C je dobrý“ je kladná
- Oblíbená metoda v kombinatorice
- Volme n dost velké a ať $k = \lceil n\kappa \rceil$
- Zvolme kód C náhodně tak, že 2^k -krát nezávisle rovnoměrně vybereme $c_i \in \{0, 1\}^n$
- Pak vybereme náhodně rovnoměrně $i \in \{1, 2, \dots, 2^k\}$
- Odešleme c_i kanálem, ten náhodně udělá chyby (další náhodnost!)
- Dekódujeme na nejbližšího souseda; jaká je pst, že dekodujeme zpátky c_i ?

- Vyberu náhodně C, i
- Kanál náhodně vybere vektor chyb $e \in \{0, 1\}^n$ (nezávisle na C, i)
- Příjmu $\tilde{c} = c_i + e$; ať $D(\tilde{c})$ je nejbližší soused \tilde{c}
- Pokud $P[D(\tilde{c}) = c_i] > 1 - \varepsilon$, tak vyhrávám
- Zádrhel: Věta tvrdí, že existuje jeden spolehlivý kód, ne „náhodný kód“
- Trik

$$P[D(\tilde{c}) = c_i] = \sum_C P[D(\tilde{c}) = c_i | \text{zvolili jsme } C] P[\text{zvolili jsme } C]$$

- Pokud pro každé C je $P[D(\tilde{c}) = c_i | \text{zvolili jsme } C] \leq 1 - \varepsilon$, tak také $P[D(\tilde{c}) = c_i] \leq 1 - \varepsilon$

$$P[D(\tilde{c}) = c_i] \geq 1 - \varepsilon$$

- Zvolím n velké, $k = \lceil n\kappa \rceil$, $\alpha > 0$ aby $p + \alpha < 1/2$
- Vyberu náhodně C, i
- Kanál náhodně vybere vektor chyb $e \in \{0, 1\}^n$ (nezávisle na C, i)
- Jaká je pravděpodobnost jevu „nastalo více než $n(p + \alpha)$ chyb“?
- Nejvýše $e^{-n\alpha^2/2}$, viz slajd 8

Chyb je nejvýše $n(p + \alpha)$

- Značme X^c jev „nastalo více než $n(p + \alpha)$ chyb“
- Víme $P[X^c] \leq e^{-n\alpha^2/2}$, viz slajd 8
- Zvolíme n, α , aby $e^{-n\alpha^2/2} < \varepsilon/2$
- Ukážeme $P[D(\tilde{c}) = c_i | X] \geq 1 - \varepsilon/2$
- Potom

$$\begin{aligned} P[D(\tilde{c}) = c_i] &= P[D(\tilde{c}) = c_i | X^c]P[X^c] + P[D(\tilde{c}) = c_i | X]P[X] \\ &\geq P[D(\tilde{c}) = c_i | X]P[X] > (1 - \varepsilon/2)(1 - \varepsilon/2) \\ &> 1 - \varepsilon \end{aligned}$$

$$P[D(\tilde{c}) = c_i | \text{málo chyb}] \geq 1 - \varepsilon/2$$

- Představujme si, že nejdřív kanál vygeneruje $e \in \{0, 1\}^n$ (s nejvýš $n(p + \alpha)$ jedničkami) ...
- ... pak zvolíme i , pak $c_i \in \{0, 1\}^n$...
- ... a až pak zvolíme zbytek kódových slov C
- Všechno jsou to nezávislé veličiny, takže na časovém pořadí nezáleží
- Z tohoto pořadí je ale vidět, že \tilde{c} je nezávislá na volbě $C \setminus \{c_i\}$
- Rozdělení $\tilde{c} = c_i + e$ je rovnoměrné na $\{0, 1\}^n$ protože c_i je rovnoměrně rozdělená (vzpomeňte si na začátek hodiny)

Pst, že $d(c_i, c_j) < n(p + \alpha)$

- Kanál udělal málo chyb, tedy \tilde{c} leží ve vzdálenosti $\leq n(p + \alpha)$ od c_i
- Jaká je pst, že pro nějaké $j \neq i$ leží c_j v kouli se středem \tilde{c} a poloměrem $n(p + \alpha)$?
- Fixujme nejdřív j . Zajímá nás

$$P[d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X]$$

- Proč? Protože

$$P[D(\tilde{c}) = c_i | X] \geq 1 - P[\exists j \neq i, d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X]$$

- Protože \tilde{c} , c_j jsou nezávislé rovnoměrně rozdělené, je

$$P[d(c_j, \tilde{c}) \leq n(p + \alpha) | c_i + e = \tilde{c}, X] \leq \frac{V(n, \lfloor n(p + \alpha) \rfloor)}{2^n}$$