

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

27. dubna 2020

- Prefixové kódy jsou fajn
- Entropie dává dolní odhad na délku kódu (bez šumu)
- Jenssenova nerovnost (chcete ji?)

- $V(n, r)$ buď objem koule v $\{0, 1\}^n$ o poloměru r
- $V(n, r) = \sum_{i=0}^r \binom{n}{i}$
- Zhruba $H(r/n, 1 - r/n) = \frac{1}{n} \log(V(n, r))$

Theorem

Bud' $n \in \mathbb{N}$. Pro všechna $0 \leq r \leq n/2$ platí

$$V(n, r) < 2^{nH(r/n, 1-r/n)}$$

- Máme

$$\begin{aligned} nH(r/n, 1 - r/n) &= -n(r/n \log(r/n) + (1 - r/n) \log(1 - r/n)) \\ &= -r \log(r/n) + (r - n) \log(1 - r/n) \end{aligned}$$

$$V(n, r) < 2^{nH(r/n, 1-r/n)}$$

- $nH(r/n, 1-r/n) = -r \log(r/n) + (r-n) \log(1-r/n)$
- Dosadíme do exponenciální funkce:

$$\begin{aligned} 2^{nH(r/n, 1-r/n)} &= 2^{-r \log(r/n) + (r-n) \log(1-r/n)} = \\ &= \left(\frac{r}{n}\right)^{-r} (1-r/n)^{r-n} = \frac{n^n}{r^r (n-r)^{n-r}} \end{aligned}$$

- To už vypadá skoro jako kombinační číslo...

$$n^n = (n-r+r)^n = \sum_{i=0}^n \binom{n}{i} r^i (n-r)^{n-i} > \sum_{i=0}^r \binom{n}{i} r^i (n-r)^{n-i}$$

- Protože $r \leq n-r$, tak $r^i (n-r)^{n-i} \geq r^r (n-r)^{n-r}$ pro $i \leq r$
- Tedy

$$n^n > \sum_{i=0}^r \binom{n}{i} r^i (n-r)^{n-i} \geq \sum_{i=0}^r \binom{n}{i} r^r (n-r)^{n-r}$$

$$V(n, r) < 2^{nH(r/n, 1-r/n)}$$

- Máme

$$2^{nH(n/r, 1-n/r)} = \frac{n^n}{r^r (n-r)^{n-r}}$$

- A zároveň

$$n^n > \sum_{i=0}^r \binom{n}{i} r^i (n-i)^{n-i}$$

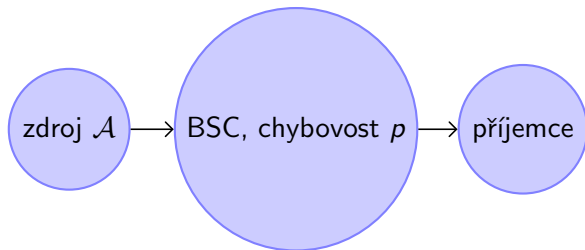
- Po vydělení druhé nerovnosti $r^r (n-r)^{n-r}$ dosazení do první rovnosti:

$$2^{nH(n/r, 1-n/r)} > \sum_{i=0}^r \binom{n}{i} = V(n, r).$$

Binární symetrický kanál

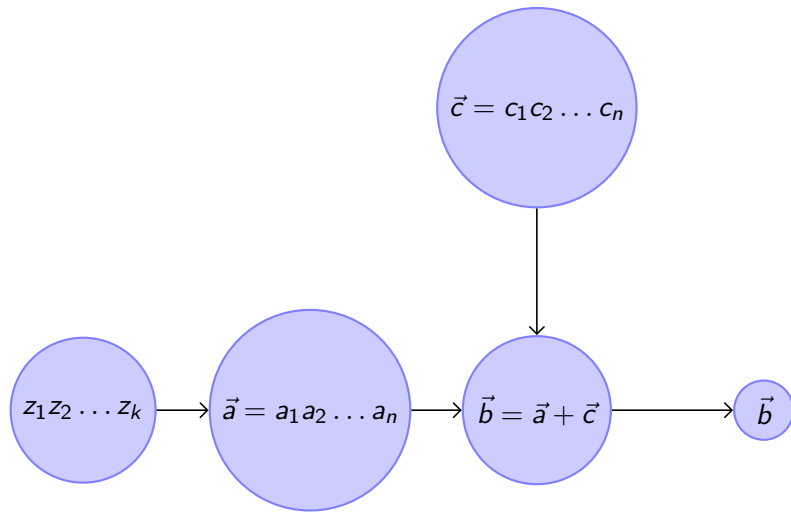
- „Náhodná funkce“ odeslané slovo \mapsto přijaté slovo
- Formálně je kanál soubor pravděpodobnostních rozdělení přijatých slov, jedno pro každé odeslané slovo
- Budeme uvažovat jenom binární symetrický kanál (BSC) bez paměti
- Parametr $p \in (0, 1)$
- Každý bit odeslané zprávy má pravděpodobnost p , tzv. **chybovost**, že se při průchodu kanálem přepne
- Přepnutí i -tého a j -tého bitu jsou nezávislé jevy (pozor: reálné kanály často dělají chyby v dávkách)

Posílání kanálem v obraze I



- Mějme BSC s chybovostí p
- Cvičení: Búno $p \leq 1/2$
- Vyšší $p \Rightarrow$ větší požadavky na opravy chyb \Rightarrow méně husté kódy
- Jak přesně ale závisí hustota kódu na chybovosti p ?
- Mějme zdroj $\mathcal{A} \rightarrow$ kód $C \rightarrow$ kanál \rightarrow dekodování
- Pro jednoduchost: \mathcal{A} ať je nad $\{0, 1\}$ a rovnoměrně rozdělený (pst $0=pst\ 1=1/2$)
- Víme, že pro $p = 0$ potřebujeme na přenos 1 znaku v průměru 1 znak kódového slova

Posílání kanálem v obraze II

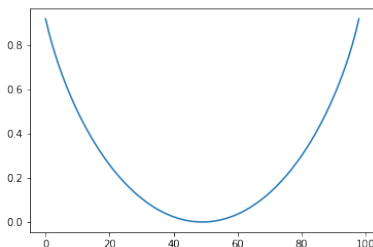


Kód, který věděl příliš mnoho

- Kanál si můžeme představovat jako informační zdroj \mathcal{B} , který s pstí p posílá $c_i = 1$, jinak $c_i = 0$...
- ... když do kanálu pošleme písmeno a_i , tak přijatý znak bude $b_i = a_i + c_i \pmod{2}$
- Pokud správně určíme odeslané ~~přijaté~~ slovo, tak zároveň určíme a_1, \dots, a_n a c_1, c_2, \dots, c_n . Tedy dekódování = příjem zprávy od **dvou** zdrojů \mathcal{A} a \mathcal{B}
- Ne úplně korektní argument: Předpokládejme, že máme blokový kód, který vždy dekóduje správně
- Pokud kódujeme bloky z \mathcal{A} po k znacích z \mathcal{A} a délka kódu je n , tak nutně $n \geq k + nH(\mathcal{B})$; to se upraví na $n(1 - H(\mathcal{B})) \geq k$
- Tedy $1 - H(\mathcal{B}) \geq k/n$

- Odvodili jsme (s dírou v argumentu), že hustota kódu je nutně $\leq 1 - H(\mathcal{B})$
- $H(\mathcal{B}) = -p \log_2 p - (1 - p) \log_2(1 - p)$
- Kapacita kanálu s chybovostí p je

$$C(p) = 1 - H(\mathcal{B}) = 1 + p \log_2 p + (1 - p) \log_2(1 - p)$$



$$C(p) = 1 - H(\mathcal{B}) = 1 + p \log_2 p + (1 - p) \log_2(1 - p)$$

- Pro $p = 0, 1$ je kanál zcela předvídatelný (kapacita 1)
- Pro $p = 1/2$ máme kapacitu 0
- Platí $C(p) = C(1 - p)$