

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

7. května 2020

Důkaz Černovovy Čebyševovy nerovnosti

Věta (Černov)

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 $\sum_{i=1}^n z_i$ leží v intervalu $[n(p - \alpha), n(p + \alpha)]$ s pravděpodobností aspoň $1 - 2e^{-n\alpha^2/2}$.

Věta (Čebyšev)

Bud' X náhodná veličina se střední hodnotou μ a rozptylem σ^2 (obojí konečné). Bud' $a > 0$. Potom $P[|X - \mu| \geq a] \leq \sigma^2/a^2$.

Pro nás $X = \sum_{i=1}^n Z_i$ má $\mu = np$ a $\sigma = np(1 - p) \leq n/4$; volíme $a = n\alpha$. Pak

$$P \left[\left| \sum_{i=1}^n Z_i - np \right| \geq n\alpha \right] \leq \frac{np(1-p)}{n^2\alpha^2} = \frac{p(1-p)}{n\alpha^2} \leq \frac{1}{4n\alpha^2}$$

Směřujeme k důkazu Čebyševa

Věta (Markov)

Bud' Y nezáporná náhodná veličina se střední hodnotou μ . Pak pro každé $a > 0$ platí

$$P[Y \geq a] \leq \mu/a.$$

- Uděláme důkaz pro Y nabývající konečně mnoha hodnot
 $0 \leq y_1 < \dots < y_m$
- Nechť pro nějaké Y a $a > 0$ je $P[Y \geq a] > \mu/a$
- značme j největší, že $y_j < a$
- Potom

$$\mu = \sum_{i=1}^m P[Y = y_i]y_i = \sum_{i=1}^j P[Y = y_i]y_i + \sum_{i=j+1}^m P[Y = y_i]y_i$$

$$\mu = \sum_{i=1}^m P[Y = y_i]y_i = \sum_{i=1}^j P[Y = y_i]y_i + \sum_{i=j+1}^m P[Y = y_i]y_i$$

- První součet je ≥ 0 , druhý je $\geq P[Y \geq a]a$
- Tedy

$$\mu \geq 0 + P[Y \geq a]a > \frac{\mu}{a} \cdot a = \mu$$

- To je spor

Věta (Markov)

Bud' Y nezáporná náhodná veličina se střední hodnotou μ . Pak pro každé $a > 0$ platí

$$P[Y \geq a] \leq \mu/a.$$

Věta (Čebyšev)

Bud' X náhodná veličina se střední hodnotou μ a rozptylem σ^2 (obojí konečné). Bud' $a > 0$. Potom $P[|X - \mu| \geq a] \leq \sigma^2/a^2$.

- Bud' X náhodná veličina; $\mu = EX$
- Trik: Uvažme náhodnou veličinu $Y = |X - \mu|^2$
- Základy pravděpodobnosti: $EY = \text{var } X = \sigma^2$
- Y je nezáporná, tedy Markov dá $P[|X - \mu|^2 \geq a^2] \leq \sigma^2/a^2$

- Mějme binární (n, k, d) -kód C
- Před časem jsem vám tvrdil, že je rozumné měřit spolehlivost kódu pomocí relativní vzdálenosti d/n
- Zkusme dokázat, že pokud p je chybovost kanálu, n je velké a třeba $d/n > 2,1p$, tak spolehlivost jde k 1
- Protože $d > 2,1pn \geq 2,09pn + 1$, tak umíme opravit aspoň $1,04pn$ chyb
- Volme $\alpha = 0,04p$. Pst, že počet chyb bude $\leq n(p + \alpha) = n(p + 0,04p) = 1,04pn$ je aspoň $1 - e^{-\alpha^2 n/2} \rightarrow 1$
- Tedy s pravděpodobností skoro 1 opravíme všechny chyby
- Drobná vada na kráse: Potřebuji n řádově $1250/p^2$

Hammingův odhad a entropie

- Hammingův odhad jsme odvodili kombinatoricky
- V logaritmické formě $n \geq k + \log_2(V_{r,n})$
- Přitom se bavíme o situaci, kdy vždy opravíme aspoň r chyb
- Alternativní důkaz pomocí entropie: Uvažme kanál, který z odeslaného slova $c \in \{0, 1\}^n$ dělá $\tilde{c} = c + e$
- Ať vektor $e \in \{0, 1\}^n$ je rovnoměrně rozložený v kouli o středu 0 a poloměru r
- Pak entropie e je $\log_2(V_{r,n})$
- Pokud kód opraví $\leq r$ chyb, tak kódujeme bezchybně
- Z \tilde{c}, c vypočteme e , tedy přijaté slovo je informační zdroj entropie $k + \log_2(V_{r,n})$
- Věta o kódování bez šumu: délka kódu musí být aspoň $k + \log_2(V_{r,n})$

Asymptotický Hammingův odhad [Kaiser 12.4]

- Pro $\delta \in [0, 1]$ značme $\alpha(\delta) = \limsup_{n \rightarrow \infty} A(n, \lceil \delta n \rceil) / n$ [sekce 12.3]
- Tedy $\alpha(\delta)$ je zhruba maximální hustota kódu o relativní vzdálenosti δ pro $n \rightarrow \infty$
- Jaký je horní odhad na $\alpha(\delta)$?
- Pro $\lceil \delta n \rceil$ liché máme z Hamminga

$$A(n, \lceil \delta n \rceil) + \log_2 \left(V_{\frac{\lceil \delta n \rceil - 1}{2}, n} \right) \leq n$$
$$\frac{A(n, \lceil \delta n \rceil)}{n} \leq 1 - \frac{\log_2 \left(V_{\frac{\lceil \delta n \rceil - 1}{2}, n} \right)}{n}$$

- Pro $\lceil \delta n \rceil$ sudé máme třeba

$$\frac{A(n, \lceil \delta n \rceil)}{n} \leq \frac{A(n, \lceil \delta n \rceil - 1)}{n} \leq 1 - \frac{\log_2 \left(V_{\frac{\lceil \delta n \rceil - 2}{2}, n} \right)}{n}$$

- Na odhad $\alpha(\delta)$ zbývá zdola odhadnout $\frac{\log_2(V_{r,n})}{n}$, kde $r \approx \delta n / 2$

Věta

Pro každé $n > k$ platí $\binom{n}{k} \geq \frac{n^n}{(n-k)^{n-k}(k+1)^{k+1}}$

- Máme

$$\ln \left(\binom{n}{k} \right) = \ln(n) + \ln(n-1) + \dots + \ln(n-k+1) - \\ - \ln(k) - \ln(k-1) - \dots - \ln(1)$$

- Pomocí integrování máme (\ln je rostoucí fce)

$$\sum_{i=1}^k \ln i \leq \int_1^{k+1} \ln x dx = [x(\ln(x) - 1)]_1^{k+1} = (k+1)(\ln(k+1) - 1) + 1$$

Odhad na $\binom{n}{k}$

- Podobně

$$\begin{aligned}\sum_{i=n-k+1}^n \ln i &\geq \int_{n-k}^n \ln x = [x(\ln(x) - 1)]_{n-k}^n \\ &= n(\ln(n) - 1) - (n-k)(\ln(n-k) - 1)\end{aligned}$$

- Tedy

$$\begin{aligned}\ln \binom{n}{k} &\geq n(\ln(n) - 1) - (n-k)(\ln(n-k) - 1) - \\ &\quad - (k+1)(\ln(k+1) - 1) - 1 \\ &\geq n \ln n - (n-k) \ln(n-k) - (k+1) \ln(k+1)\end{aligned}$$

- Odlogaritmujeme:

$$\binom{n}{k} \geq \frac{n^n}{(n-k)^{n-k}(k+1)^{k+1}}$$

Odhad na $\log_2(V_{r,n})/n$

- Upravme nerovnost z předchozího slajdu na

$$\binom{n}{k} \geq \frac{n^n}{(n-k)^{n-k}(k+1)^{k+1}} = \frac{1}{(1-k/n)^{n-k}(k/n+1/n)^k(k+1)}$$

- Máme

$$\log_2(V_{r,n}) \geq \log_2 \left(\binom{n}{r} \right) \geq \\ -(n-r) \log_2(1-r/n) - r \log_2(r/n+1/n) - \log_2(r+1)$$

- Po vydělení n máme dolní odhad na $\log_2(V_{r,n})/n$:

$$-(1-\frac{r}{n}) \log_2(1-\frac{r}{n}) - \frac{r}{n} \log_2(\frac{r}{n} + \frac{1}{n}) - \frac{\log_2(r+1)}{n}$$

- To vypadá skoro jako $H(r/n, 1-r/n) \dots$

$\log_2(V_{r,n})/n$ vs. $H(r/n)$

- Značme $\chi(r, n)$ rozdíl

$$-(1 - r/n) \log_2(1 - r/n) - r/n \log_2(r/n + 1/n) - \frac{\log_2(r + 1)}{n} - H(r/n)$$

- Máme $\chi(r, n) = -r/n(\log_2(r/n + 1/n) - \log_2(r/n)) - \frac{\log_2(r+1)}{n}$
- Dopsáno po přednášce: Protože \log_2 je konkávní, tak pro $x > 0$ je

$$\log_2(x + \frac{1}{n}) \geq \log_2(x) + \log'_2(x) \cdot \frac{1}{n} = \log_2(x) + \frac{1}{\ln(2)xn}$$

- Pro $r < n$ a $n \rightarrow \infty$ jde $\chi(r, n)$ k 0, třeba protože (je $r < n$):

$$|\chi(r, n)| \leq \frac{1}{\ln(2)n} + \log_2(n)/n < \frac{3}{\sqrt{n}}$$

- Tedy $\log_2(V_{r,n})/n \geq H(r/n) + \chi(r, n)$ pro $\chi(r, n) \rightarrow 0$

Asymptotický Hamming

- Spočetli jsme

$$\frac{A(n, \lceil \delta n \rceil)}{n} \leq 1 - \frac{\log_2 \left(V_{\lfloor \frac{\lceil \delta n \rceil - 1}{2} \rfloor, n} \right)}{n}$$

- Zvolme pevné $\delta > 0$, značme $r(n) = \lfloor \frac{\lceil \delta n \rceil - 1}{2} \rfloor$

- Přitom

$$\frac{\log_2 (V_{r(n), n})}{n} \geq H(r(n)/n) + \chi(r(n), n)$$

- Tedy

$$\frac{A(n, \lceil \delta n \rceil)}{n} \leq 1 - H(r(n)/n) - \chi(r(n)/n, n)$$

- Nechme $\delta > 0$ konstantní a pošleme $n \rightarrow \infty$.
- Protože $r(n)$ je $(\lceil \delta n \rceil - 1)/2$ nebo $(\lceil \delta n \rceil - 2)/2$, tak $r(n)/n \rightarrow \delta/2$
- Pak $\chi(r(n)/n, n) \rightarrow 0$
- Dále $H(r(n)/n) \rightarrow H(\frac{\delta}{2})$
- Potom

$$\alpha(\delta) = \limsup_n \frac{A(n, \lceil \delta n \rceil)}{n} \leq 1 - H\left(\frac{\delta}{2}\right)$$

- Příklad $\delta = 0$ musíme udělat zvlášť: je $A(n, 0)/n = n/n = 1 - H(0)$
- Pro zájemce: Gilbert–Varshamův odhad [Kaiser 12.3] říká, že pro $\delta \in [0, 1/2]$ je

$$\alpha(\delta) \geq 1 - H(\delta).$$

Existuje rodina kódů s $k/n, d/n$ odraženými od 0?

- Náhodně generovat množinu velikosti $2^{\kappa n}$ není moc praktické
- Existuje nějaký typ kódů, kde lze $n \rightarrow \infty$, aby k/n ani d/n nešlo k nule (takovým množinám kódů se říká dobré kódy)
- Existují: Justesenovy kódy (viz Kasier 12.5)

- Pravděpodobnostní techniky (NTIN022)
- Konvoluční kódy (libovolně velké bloky)
https://en.wikipedia.org/wiki/Convolutional_code
- LDPC kódy, Turbo kódy (hustota zhruba rovná kapacitě kanálu)
https://en.wikipedia.org/wiki/Low-density_parity-check_code
https://en.wikipedia.org/wiki/Turbo_code
- Statistická fyzika (NUFY094?)
- Vzájemná informace a důsledky pro šifrování (jako na přednášce Kryptografické systémy NMMB201)
- Zpracování signálu (NPFL109)