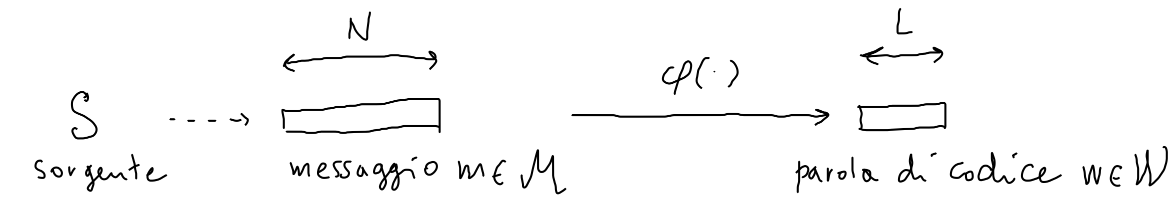
TASSO DI UN CODICE



Messaggi $m \in \mathcal{M}$: probabilità $q(m)$, lunghezza $n(m)$ $\Rightarrow$ valore atteso della lunghezza di un messaggio

$$E[N] = \sum_{m \in \mathcal{M}} q(m) \cdot n(m)$$

Parole di codice $w \in \mathcal{W}$: probabilità $p(w)$, lunghezza $\ell(w)$ $\Rightarrow$ valore atteso della lunghezza di una parola di codice

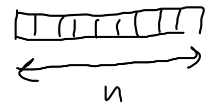
$$E[L] = \sum_{w \in \mathcal{W}} p(w) \cdot \ell(w)$$

Tasso del codice è il rapporto $R \triangleq \dfrac{E[L]}{E[N]}$; $\dfrac{1}{R}$ è il rapporto di compressione

Nel caso di codici B-B o B-LV , $R = \dfrac{E[L]}{n}$ dove n è la lunghezza dei blocchi della sorgente. (alfabeto primario)

Codifica di sorgente: Data una sorgente, individuare un codice $C=(\mathcal{M},\varphi,W)$ che sia u.d. e a tasso minimo (o approssimativamente tale).

Codici B-LV. $\mathcal{M}=\mathcal{A}^n$



Iniziamo da <u>n = 1</u>. In questo caso $R=\dfrac{E[L]}{n}=E[L]$.

Se $\mathcal{A}=\{a_1,a_2,\dots,a_K\}$ prob.: $p_1,p_2,\dots,p_K$

$W=\{w_1,w_2,\dots,w_K\}$ lunghezze $l_1,l_2,\dots,l_K$

$$R = E[L]=\sum_{i=1}^{K} p_i l_i$$

Teo (3.13) (Teorema di Shannon). Se S è una sorgente (stazionaria e senza memoria) con d.p. $\vec{p}=(p_1,p_2,\dots,p_K)$ la lunghezza medie delle parole di qualunque codice D-ario u.d. è almeno pari all'entropia (D-aria) della sorgente: $E[L]\geq H_D(\vec{p})=-\sum_{i=1}^{K}p_i\log_D p_i$

Dim. Consideriamo la d.p. $\vec{q} = (q_1, \dots, q_k)$ con $q_i \doteq \dfrac{D^{-l_i}}{\sum_{j=1}^{k} D^{-l_j}} = \dfrac{D^{-l_i}}{\alpha}$

dove $\alpha = \sum_{j=1}^{k} D^{-l_j}$. Scriviamo la divergenza informazionale tra $\vec{p}$ e $\vec{q}$:

Div. inform.

$$0 \le D(\vec{p} \| \vec{q}) = \sum_{i=1}^{k} p_i \log_D\left[\frac{p_i}{D^{-l_i}} \alpha\right] = \sum_{i=1}^{k} p_i \log_D p_i + \sum_{i=1}^{k} p_i \overbrace{\log_D(D^{l_i})}^{l_i} + \underbrace{\sum_{i=1}^{k} p_i}_{1} \log_D \alpha$$

$\log_D x = \dfrac{\log_2 x}{\log_2 D}$

$$= \sum_{i=1}^{k} p_i \log_D p_i + \underbrace{\sum_{i=1}^{k} p_i l_i}_{E[L]} + \log_D \alpha$$

$$\Rightarrow E[L] \ge \underbrace{-\sum_{i=1}^{k} p_i \log_D p_i}_{H_D(\vec{p})} \overbrace{- \underbrace{\log_D \alpha}_{\alpha \le 1}}^{\ge 0} \ge H_D(\vec{p}). \quad \text{QED}$$

$\sum_{j=1}^{k} D^{-l_j} \le 1$ (disuguaglianza di McMillan-Kraft).

Si ha uguaglianza $(E[L] = H_D(\vec{p}))$ sse $\alpha = 1$ e $\vec{p} = \vec{q}$, quindi $p_i = D^{-l_i}/\alpha = D^{-l_i}$.

(distribuz. diadica)

Se $\vec{p}$ è uniforme: $p_i = 1/K$, $i = 1 \dots K$ $\rightarrow$ $E[L] \geq H_D(\vec{p}) = \log_D K$

è la lunghezza in D-it delle lettere primarie
$\rightarrow$ non riesco ad avere compressione

La differenza

$$r \stackrel{\Delta}{=} E[L] - H_D(\vec{p}) \quad (\geq 0)$$

è detta ridondanza del codice.

Esistono codici con tasso prossimo all'entropia? (ridondanza prossima a zero?)

Se riuscissi ad avere $l_i \approx -\log_D p_i$, allora $E[L] = \sum_{i=1}^{K} p_i l_i \approx -\sum_i p_i \log_D p_i = H_D(\vec{p})$

Un approccio è porre $l_i = \lceil -\log_D p_i \rceil$ (codice di Shannon-Fano).

Teo (3.14). Se S è una sorgente con d.p. $\vec{p}$, esiste un codice D-ario a prefisso tale che
$$E[L] < H_D(\vec{p}) + 1.$$

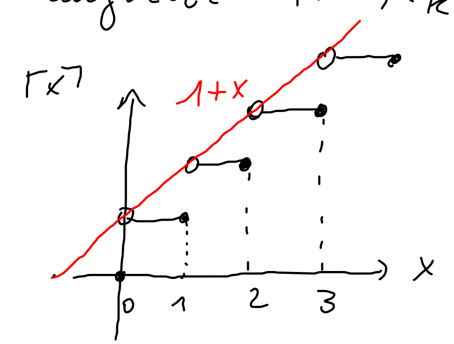
Teo (3.14). Se $\mathcal{S}$ è una sorgente con d.p. $\vec{p}$, esiste un codice D-ario a prefisso tale che

$$E[L] < H_D(\vec{p}) + 1 .$$

Dim. $l_i = \lceil -\log_D p_i \rceil$, $i = 1 \dots K$. Tali lunghezze soddisfano McMillan-Kraft:

infatti $D^{-l_i} = D^{-\lceil -\log_D p_i \rceil} \leq p_i$, $i = 1 \dots K$

$\Rightarrow \sum_{i=1}^{K} D^{-l_i} \leq \sum_{i=1}^{K} p_i = 1$. $\Rightarrow$ esiste codice a prefisso con lunghezze $l_1, \dots, l_K$.



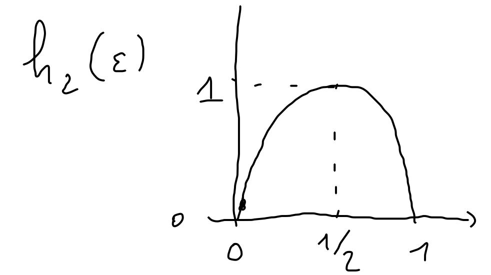
Ma $l_i = \lceil -\log_D p_i \rceil < 1 - \log_D p_i$

$$E[L] = \sum_{i=1}^{K} p_i l_i < \sum_i (-\log_D p_i + 1) \cdot p_i$$

$$= \underbrace{-\sum_{i=1}^{K} p_i \log_D p_i}_{= H_D(\vec{p})} + \underbrace{\sum_{i=1}^{K} p_i}_{= 1} = H_D(\vec{p}) + 1 . \quad \text{QED}$$

Esistono casi in cui $E[L]$ è arbitrariamente vicina a $H_D(\vec{p})+1$:
(per Shannon-Fano)

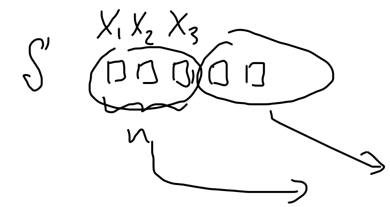
Esempio : $\mathcal{A}=\{0,1\}$, $\vec{p}=(\varepsilon, 1-\varepsilon)$, per qualche $\varepsilon>0$.



La lunghezza media $E[L]$ è comunque 1 (o ≥ 1).

$$H_D(\vec{p}) = -\varepsilon \log_D \varepsilon - (1-\varepsilon)\log_D(1-\varepsilon) \xrightarrow{\varepsilon\to 0} 0$$

quindi $H_D(\vec{p})+1 - \underbrace{E[L]}_{=1} \xrightarrow[\varepsilon\to 0]{} 0$



Per migliorare, prendiamo blocchi di n simboli della sorgente: $(n>1)$
(sorgente estesa $\mathcal{S}^n$). Ricordiamo che $H_D(\mathcal{S}^n) = H_D(X_1,\dots,X_n) = n\cdot H_D(X_1) = n H_D(\mathcal{S})$.

$$nH_D(\mathcal{S}) = H_D(\mathcal{S}^n) \leq E[L^{(n)}] < H_D(\mathcal{S}^n)+1 = nH_D(\mathcal{S})+1 \Rightarrow H_D(\mathcal{S}) \leq \frac{E[L^{(n)}]}{n} < H_D(\mathcal{S}) + \frac{1}{n}$$

lungh. di una parola di codice che codifica un blocco di n simboli

Tasso R_n

⟹ Teo (3.15) (Teorema di Shannon per sorgenti estese).

Sia S una sorgente. Esiste una successione di codici D-ari per S^n con tasso $R_n = \frac{E[L^{(n)}]}{n}$ che tende, per $n \to \infty$, all'entropia della sorgente $H_D(S)$.

$$\begin{array}{ccccc} H_D(S) & \le & R_n & < & H_D(S) + 1/n \\ \downarrow & & \downarrow & & \downarrow \\ H_D(S) & & H_D(S) & & H_D(S) \end{array}$$

Shannon-Fano è un codice "asintoticamente" ottimo

—

Per $\underline{n \text{ finito}}$, il codice di Shannon-Fano non è necessariamente il codice a tasso minimo.

B - B