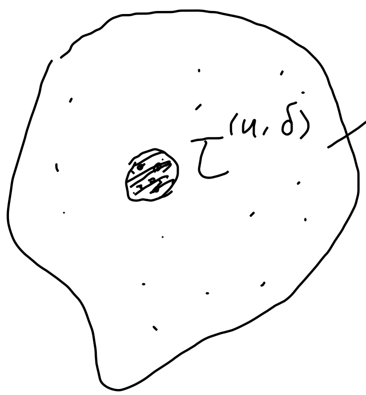
Sorgente

$\mathcal{X}^n$



seq. tipiche

$H(S) < \log|\mathcal{X}| - \delta$

Principio di equipartizionamento asintotico $(n \to \infty)$

$p(x) \approx 2^{-nH(S)}$ per ogni $x \in \mathcal{T}^{(n,\delta)}$

$|\mathcal{T}^{(n,\delta)}| \approx 2^{nH(S)}$

$Pr[X^n \in \mathcal{T}^{(n,\delta)}] \approx \underset{(1-\varepsilon)}{1}$

Esempio. $\mathcal{X} = \{0,1\}$

$n = 100$

$|\mathcal{X}^n| = 2^{100}$

$$\left.\begin{array}{c} \overbrace{00\cdots0}^{100} \\ \vdots \\ 11\cdots1 \end{array}\right\} \mathcal{X}^n$$

$\delta = 0.061$

$p = (\underset{0}{0.89}, \underset{1}{0.11}) \to H(S) \cong 0.5$

$x \in \mathcal{T}^{(n,\delta)}$ se

$$\boxed{2^{n(H(S)+\delta)} \le p(x) \le 2^{-n(H(S)-\delta)}}$$

$p(x) \approx 2^{-nH(S)} = 2^{-50} \approx 10^{-15}$

$10^{-17} \le p(x) \le 7 \cdot 10^{-14}$

$$\Pr[X^n \text{ abbia } j \text{ zeri}] = \binom{n}{j} 0.89^j \, 0.11^{n-j}$$

$\{0,1\}$

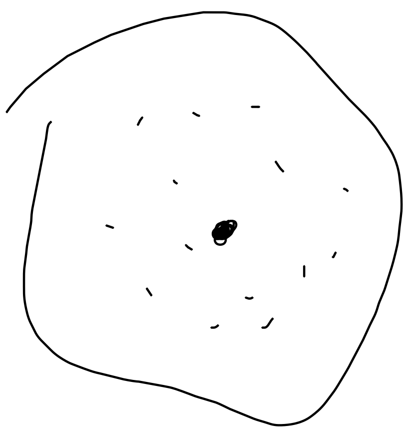
$|1|0|1|1|0|1|0|1| \rightarrow$ 3 zeri, 5 uni

j zeri

"tipo" $(3,5)$

$$\frac{|T^{(n,\delta)}|}{|\mathcal{X}^n|} \cong \frac{10^{15}}{2^{100}} = \frac{10^{15}}{10^{30}} = 10^{-15}$$

$$\Pr[X^n \in T^{(n,\delta)}] = 0.57 \rightarrow 57\% \text{ di prob. che } X^n \in T^{(n,\delta)}$$



— Tipiche in probabilità

— Tipiche in composizione

"Tipo" di una sequenza:

$\#x_1 \; \#x_2 \; \dots, \#x_K$

$(n_1, n_2, \dots, n_K) \in \mathbb{N}^K$ dove

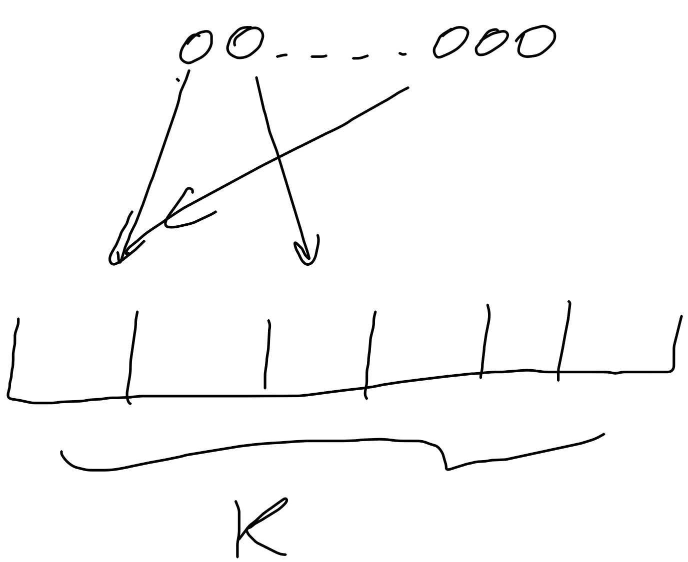
$K = |\mathcal{X}|$

n_i = numero di occorrenze del simbolo x_i nella sequenza

Numero di sequenze con la stessa composizione di una sequenza di tipo $(n_1, n_2, \dots, n_K)$:

$$\binom{n}{n_1, n_2, \dots, n_K} = \frac{n!}{n_1!\, n_2! \cdots n_K!}$$

= Numero di modi di distribuire n oggetti distinti in K contenitori



con n_1 oggetti nel contenitore 1,

$\vdots$

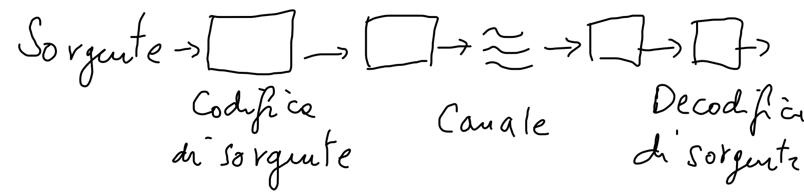
n_j oggetti nel contenitore j,

$\vdots$

n_K oggetti nel contenitore K

Codifica di sorgente

Alfabeto $\mathcal{A}$ $\mathcal{B}$



Alfabeto della sorgente (primario)

Alfabeto di codifica (secondario)

$\mathcal{A} = \{a_1, a_2, \dots, a_K\}$

$\mathcal{B} = \{b_1, \dots, b_D\}$

In pratica, spesso $D = 2$ (codifica binaria)

$\mathcal{A}^+$ ins. delle sequenze finite sull'alfabeto $\mathcal{A}$ $\xrightarrow[\text{Codifica}]{}$ $\mathcal{B}^+$ ins. delle seq. finite sull'alfabeto $\mathcal{B}$

$a_2 a_1 a_1 a_3 \in \mathcal{A}^+$

- Il codice deve essere univocamente decodificabile (u.d.) ovvero la codifica deve essere invertibile
- Possibilmente, il rapporto tra la lunghezza media della seq. secondaria e la lungh. media della seq. primaria dovrebbe essere basso, per efficienza

$*$: operatore di concatenazione

$\alpha_1, \alpha_2 \in \mathcal{A}^+ \rightarrow \alpha_1 * \alpha_2 \in \mathcal{A}^+$ concatenazione di α_1 e α_2

Codice (astratto) : funzione $\varphi : A^+ \longrightarrow B^+$ che sia
(funzione di codifica)

un omomorfismo

$$\varphi(\alpha_1 * \alpha_2) = \varphi(\alpha_1) * \varphi(\alpha_2)$$

Insieme finito $\mathcal{M} = \{m_1, m_2, \dots, m_T\} \subseteq A^+$ detto insieme dei <u>messaggi</u>

(se $\mathcal{M} = A$, allora i messaggi sono i simboli stessi dell'alfabeto)

$$\mathcal{W} = \{\varphi(m_1), \varphi(m_2), \dots, \varphi(m_T)\} \subseteq B^+$$

detto insieme delle <u>parole di codice</u>

$A = \{a, b, c\}$

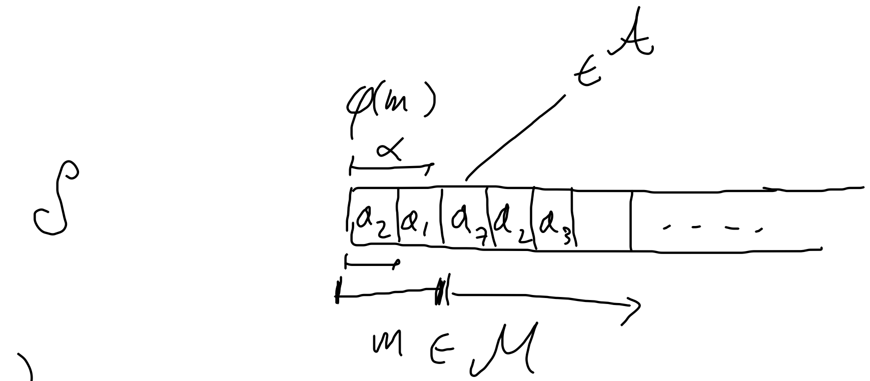
$B = \{0, 1\}$

Def. Un codice è una terna $C = (\underbrace{\mathcal{M}, \varphi}, \mathcal{W})$

A^+: abbca, aaa, bacca, ... $\} \mathcal{M}$

$\mathcal{M}$ deve permettere di segmentare qualunque $\alpha \in \mathcal{A}^+$ emessa dalla sorgente:

Def. $\mathcal{M} = \{m_1, \dots, m_T\}$ è detto <u>esauriente</u> se per ogni sequenza semi-infinita a destra su $\mathcal{A}$, esiste almeno un prefisso $\alpha \in \mathcal{M}$.



$\mathcal{M} = \{ a_2 a_1, \dots, \}$

$\mathcal{M} = \{ a_2, \dots, \}$

$\mathcal{M} = \{ a_2 a_1 a_7, \dots, \}$

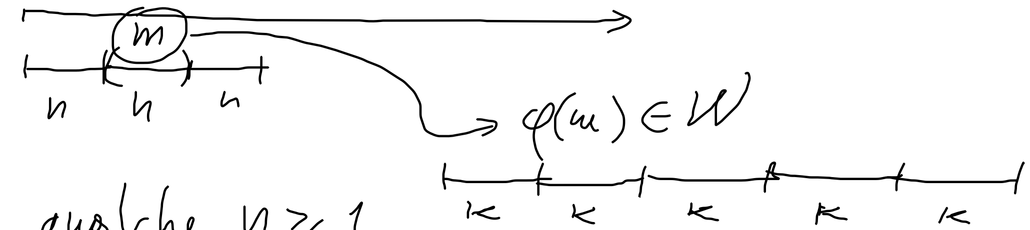
⋮

Def. Il codice $C = (\mathcal{M}, \varphi, \mathcal{W})$ è detto <u>univocamente decodificabile</u> (u.d.). se

$$\alpha \neq \alpha' \implies \varphi(\alpha) \neq \varphi(\alpha')$$
$$\forall \alpha, \alpha' \in \mathcal{A}^+$$

(φ deve essere iniettiva).

Tipi di codice di sorgente:



- Blocco-blocco (B-B): $M = A^n$ per qualche $n \geq 1$

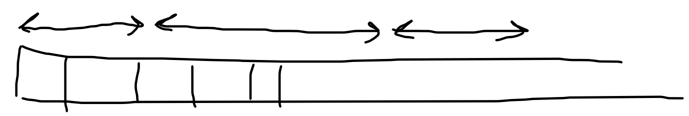
  e gli elementi di W (parole di codice) hanno tutti la stessa lunghezza (k)

  (esaurienza è garantita)

- Blocco-lunghezza variabile (B-LV): $M = A^n$ per qualche $n \geq 1$

  Gli elementi di W hanno lunghezza variabile

  (esaurienza è garantita)



- Lungh-variabile-blocco (LV-B): i Messaggi hanno lunghezza variabile;

  le parole di codice hanno lunghezza costante

- LV-LV: Sia i messaggi che le parole di codice hanno lungh. variabile

Esempi. $\mathcal{A} = \{a, b, c\}$, $\mathcal{B} = \{0, 1\}$

Seq. emessa dalla sorgente : $\alpha = bbacab$

Esempio di Codifica B-B. $\mathcal{M} = \mathcal{A} = \{a, b, c\}$, $W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{\overset{\varphi(a)}{00}, \overset{\varphi(b)}{01}, \overset{\varphi(c)}{10}\}$

$$\varphi(\alpha) = \varphi(b)\varphi(b)\varphi(a)\varphi(c)\varphi(a)\varphi(b) = 01/01/00/10/00/01 = 010100100001$$

B-LV. $\mathcal{M} = \mathcal{A}$, $W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 10, 11\}$

$b\,b\,a\,c\,a\,b$ $\quad \varphi(\alpha) = 10/10/0/11/0/10 = 1010011010$

LV-B. $\mathcal{M} = \{aa, ab, ac, b, c\}$, $W = \{\varphi(aa), \varphi(ab), \varphi(ac), \varphi(b), \varphi(c)\}$

$= \{000, 001, 010, 011, 100\}$

$$\varphi(\alpha) = \varphi(bbacab) = \varphi(b)\varphi(b)\varphi(ac)\varphi(ab) = 011/011/010/001 = 011011010001$$

$b/b/ac/ab$

LV-LV. Sia $\mathcal{M}$ come sopra , $W = \{\overset{\varphi(aa)}{000}, \overset{\varphi(ab)}{001}, \overset{\varphi(ac)}{01}, \overset{\varphi(b)}{10}, \overset{\varphi(c)}{11}\}$ $\quad \alpha : b/b/ac/ab$

$\varphi(\alpha) = 10/10/01/001$

(caso B-B e LV-B)

Per codici con parole di codice a lungh. costante, per avere univoca decodificabilità è necessario e sufficiente che tutte le parole di codice siano distinte.

$$\begin{array}{lccc} \mathcal{M} & m_1, & \dots, & m_T \\ \varphi & \downarrow\uparrow & & \downarrow\uparrow \\ W & \varphi(\underline{m}_1), & \dots, & \varphi(\underline{m}_T) \end{array}$$

$|W| = |\mathcal{M}|$

$\varphi(\alpha) = \varphi(m^{(1)})\,\varphi(m^{(2)})\,\dots$

$\downarrow m^{(1)} \quad \downarrow m^{(2)}$

Per codici con parole di codice a lunghezza variabile, ciò non è vero!

(Ambiguità strutturale)

Esempio. $\mathcal{M} = \mathcal{A} = \{a, b, c\}$

$W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 01, 001\}$

Se ricevo 000101, come la decodifico?

$0/0/01/01 \longrightarrow \underline{aabb}$

$0/001/01 \longrightarrow \underline{acb}$

???