



# Teoria dell'Informazione e Applicazioni – a.a. 2014-2015

Esercizi su **Codifica** aritmetica e compressione

11-12-2014  
Ing. P. Fazio

## CODIFICA ARITMETICA

← TAG

(24)

- ASSOCIA UN NUMERO REALE  $\in [0, 1)$  IN MANIERA UNIVOCAMENTE ALLA PAROLA DATI (MESSAGGIO) DA TRASMETTERE;
- SI BASA SUL CONCETTO DI PDF E CDF DEI SIMBOLI DELL'ALFABETO;

- SI DIVIDE RICORRENTEMENTE L'INTERVALLO  $[0, 1)$  SECONDO LE PROBABILITÀ ORIGINALI  
→ UN QUALSIASI NUMERO DELL'ULTIMO INTERVALLO OTTENUTO RAPPRESENTA UNIVOCAMENTE LA STRINGA ORIGINARIA (IN GENESIS SI POSSONO L'ESTREMO INF, SUP O LA MEDIA);

- IL NUMERO FINALE OTTENUTO RAPPRESENTA LA CDF DELLA SEQUENZA DI SIMBOLI ORIGINALI

- IN GENESIS SI USA UNA V.A. PER MAPPARE LE OCCORRENZE DEI SIMBOLI:  
DATO L'ALFABETO  $A = \{a_1, \dots, a_m\}$  allora  $X(a_i) = i$   $a_i \in A$  PUÒ ESSERE UN  
GOOD MAPPING, CON:

- PDF  $\Rightarrow P(X=i) = P(a_i)$

- CDF  $\Rightarrow F_X(i) = \sum_{k=1}^i P(X=k)$

$$\begin{aligned} \min [F_X(i)] &= 0 & F_X \text{ ripetibile} \\ \max [F_X(i)] &= 1 & \text{limit. } [0, 1) \end{aligned}$$

- SI PARTE DALLA PDF (SE IPOTIZZATA), SI COSTRUISCE LA CDF SIN  
 BASI ALLE OCCORRENZE SUCCESSIVE SI VA A ~~CONSTRUIRE~~ <sup>RESTRINGENDO</sup> L'INTERVALLO.

SI PARTE DA  $l^{(0)} = 0$ ,  $u^{(0)} = 1$  E AD OGNI PASSO INTERNUMO AVANZO:

$$l^{(m)} = l^{(m-1)} + [u^{(m-1)} - l^{(m-1)}] \cdot F_X(x_{m-1}) \quad (F_0 \equiv 0 \text{ ALL' INIZIO})$$

$$u^{(m)} = l^{(m-1)} + [u^{(m-1)} - l^{(m-1)}] \cdot F_X(x_m)$$

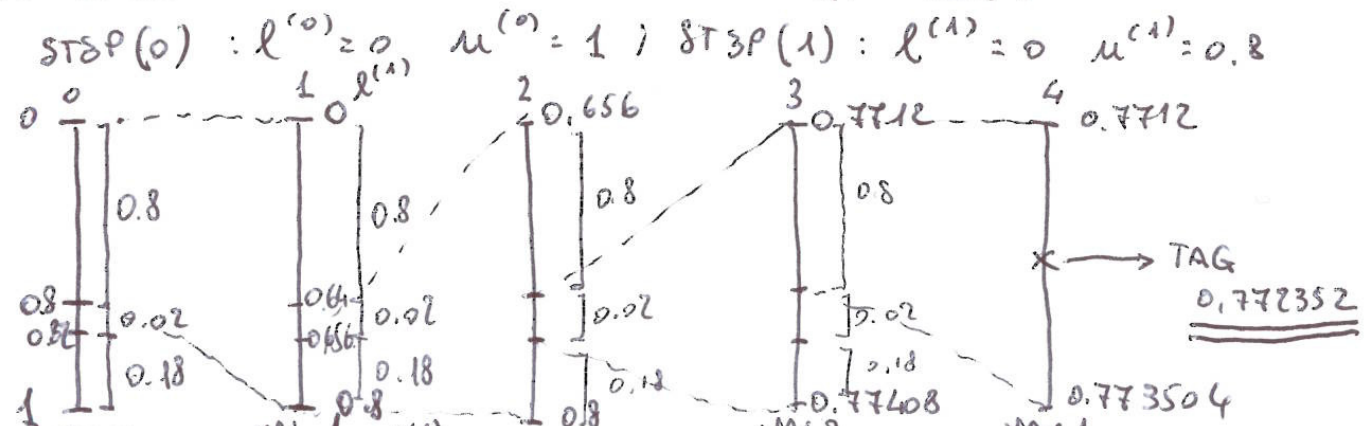
SEGUENDO IL VALORE MEDIO COME TAG, ESSO SARÀ:  $TAG(x) = \frac{u^{(m)} + l^{(m)}}{2}$

ESEMPIO 1)

$A = \{1, 2, 3\}$  con  $P(1) = 0,8$ ,  $P(2) = 0,02$ ,  $P(3) = 0,18$ , VOGLIAMO

CALCOLARE IL CODICE ARITMETICO PER LA STRINGA 1321:

$x$	$f_x$	$F_x$
1	0,8	0,8
2	0,02	0,82
3	0,18	1
//	//	//



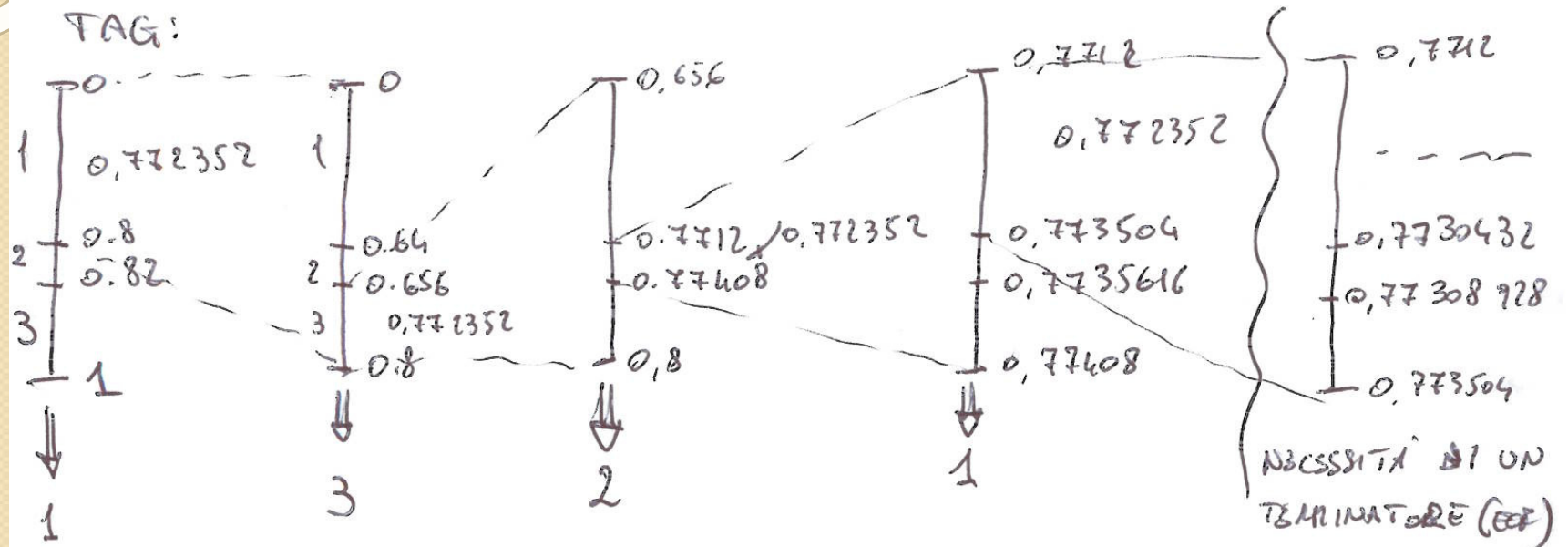


STEP 2:  $l^{(2)} = 0,656$   $u^{(2)} = 0,8$  ; STEP 3:  $l^{(3)} = 0,7712$   $u^{(3)} = 0,77408$  25  
 STEP 4:  $l^{(4)} = 0,7712$   $u^{(4)} = 0,773504$  TAG(1321) =  $\frac{0,773504 + 0,7712}{2} = \underline{\underline{0,772352}}$

- DECODIFICA:

- SI SCEGLIE IL SIMBOLO A CUI È ASSOCIATO L'INTERVALLO CHE CONTIENE IL

TAG:



LA LUNGHEZZA MEDIA È, GENERALMENTE, SUPERIORE A QUELLA DI HUFFMAN, MA È PIÙ EFFICIENTE DA UN PUNTO DI VISTA COMPUTAZIONALE. IL PRIMO TIPO, PERÒ, IN QUESTA COMPLESSA NON POTSSEN PRODOTTA IN USCITA FINCHÈ NON TERMINA IL PROCESSO.

## ESSEMPIO 2)

COMPLICAN ARITMETICAMENTE LA STRINGA "TELESTATICA".

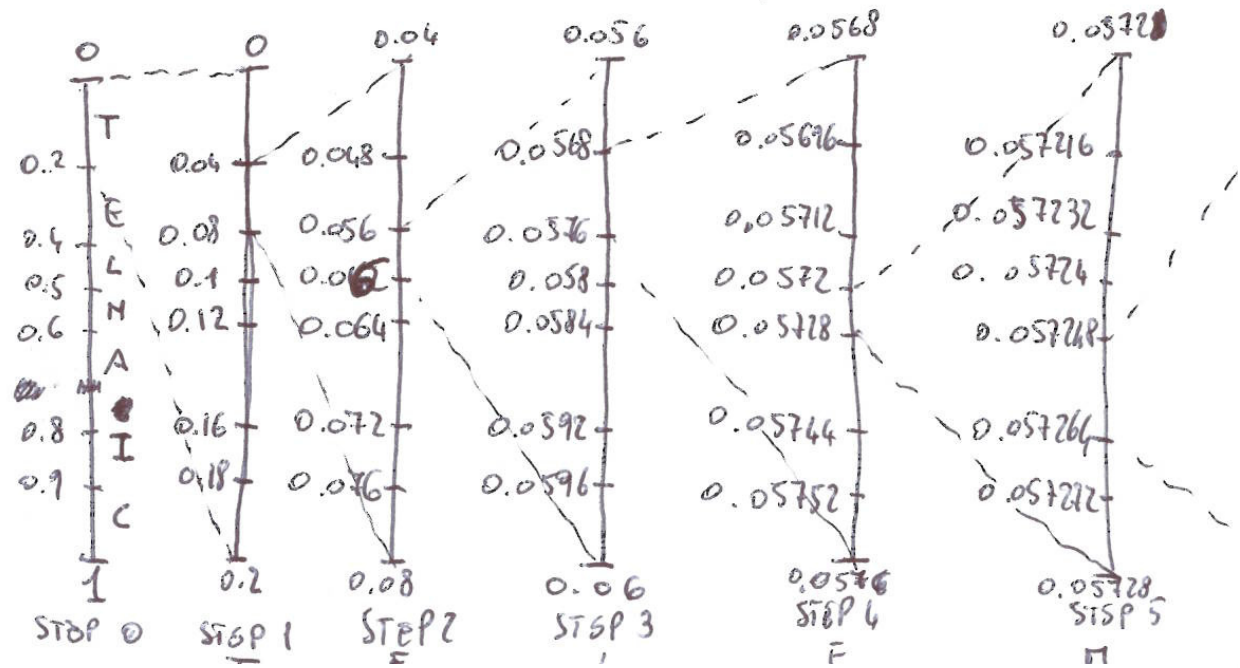
- DEFINIZIONE ALFABETO:  $\{T, E, L, M, A, I, C\}$

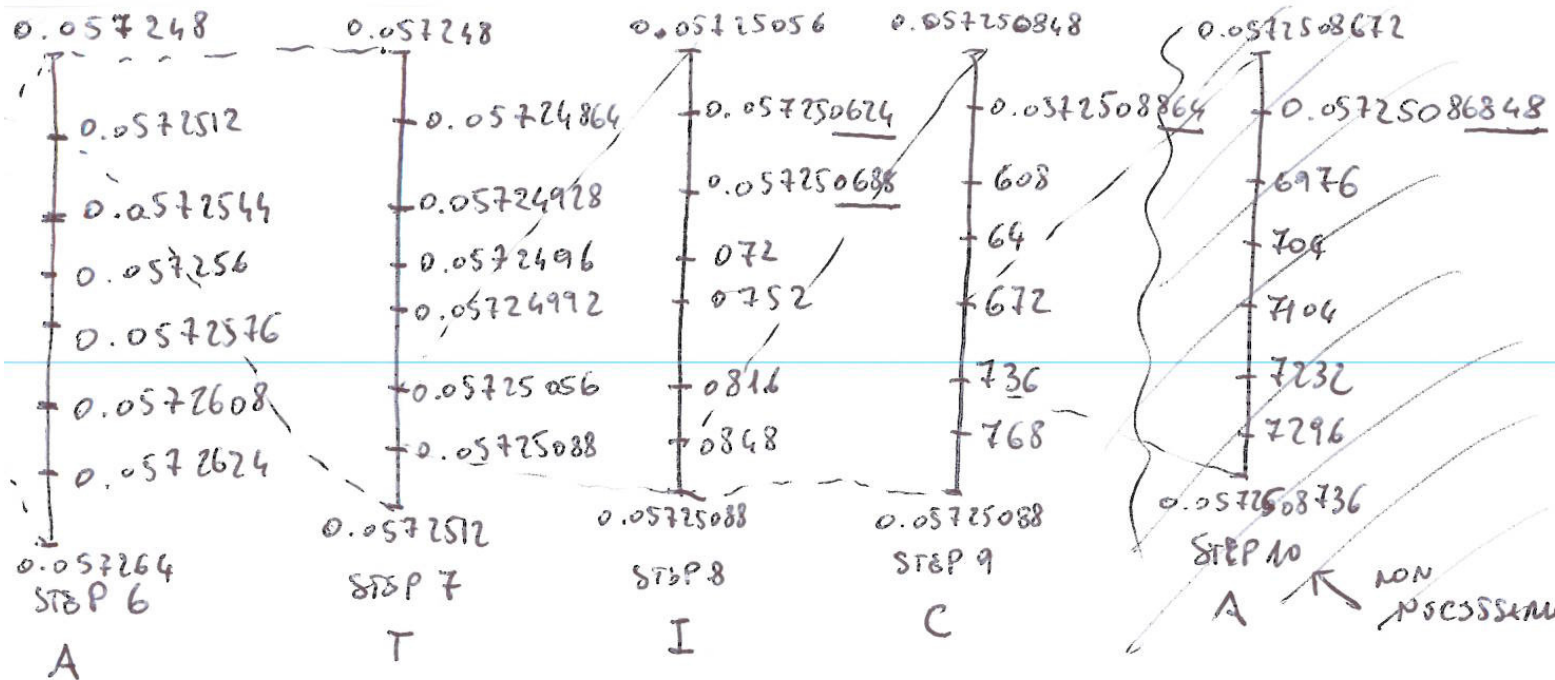
- DEFINIZIONE PDF:  $\{0.2, 0.2, 0.1, 0.1, 0.2, 0.1, 0.1\}$

- DEFINIZIONE CDF:  $\{0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 1\}$

X	P(X)	F(X)
T	0.2	0.2
E	0.2	0.4
L	0.1	0.5
M	0.1	0.6
A	0.2	0.8
I	0.1	0.9
C	0.1	1

// // //





$$TAG = \frac{0.0572508736 + 0.0572508672}{2} = \underline{\underline{0.0572508704}}$$

SI OSSERVA CHE ALL'AUMENTARE DELLA LUNGHEZZA DEL FUSCICINO SI COMPLICANO LA PRECISIONE DEL RISULTATO AUTENTA; SI RICORRE ALLA COMPLESSA ANALITICA INCREMENTALE, ACCANTONANDO LE CIFRE CHE NON VARIANO.



VEDIAMO COME COMPLETARE IN BINARIO IL TAG TROVATO (METODO DELLE MOLTIPLICAZ. SUCCESSIVE):

$$0.0572508704 \times 2 = 0.1145017408 \quad (0)$$

$$0.1145017408 \times 2 = 0.2290034816 \quad (0)$$

$$0.2290034816 \times 2 = 0.4580069632 \quad (0)$$

$$0.4580069632 \times 2 = 0.9160139264 \quad (0)$$

$$0.9160139264 \times 2 = 1.8320278528 \quad (1)$$

$$0.8320278528 \times 2 = 1.6640557056 \quad (1)$$

$$0.6640557056 \times 2 = 1.3281114112 \quad (1)$$

$$0.3281114112 \times 2 = 0.6562228224 \quad (0)$$

$$0.6562228224 \times 2 = 1.3124456448 \quad (1) \dots$$

$$\text{TAG (BIN)} = 0. \underbrace{000011101}_{\text{(APPROXIM.)}}$$

LA RICONVERSIONE AVVIENE CON:

$$\begin{aligned} & 0.2^{-1} + 0.2^{-2} + 0.2^{-3} + 0.2^{-4} + 1.2^{-5} + 1.2^{-6} + 1.2^{-7} + 0.2^{-8} + 2^{-9} \\ & = 0.03125 + 0.015625 + 0.0078125 + 0.001953125 = 0.056640625 \end{aligned}$$