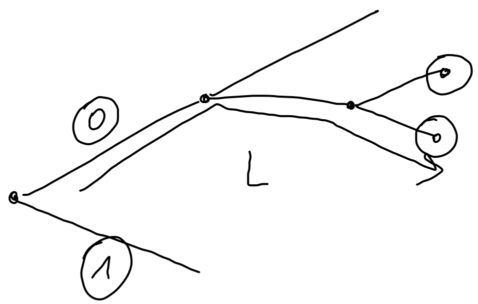
① Gioco "Twenty Questions"

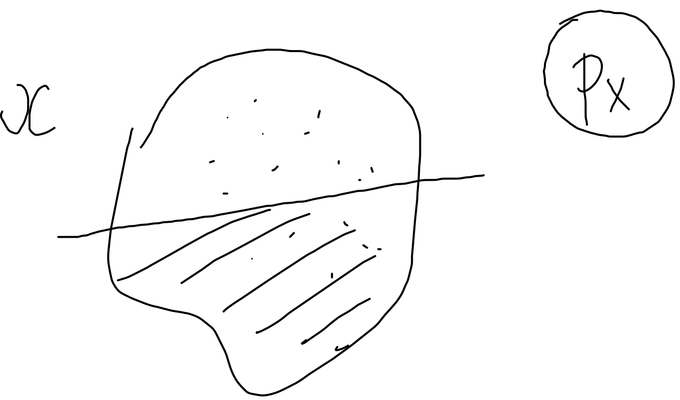
Gioco: giocatore A sceglie un elemento X da un insieme universo $\mathcal{X}$

giocatore B cerca di indovinare X attraverso domande a risposta si/no?



Supponiamo che B usi una codifica a lunghezza attesa minima (rispetto alla distribuzione p_X). $\mathbb{E}[L]$

Se osserviamo una media di 20 domande, qual è il numero minimo di elementi dell'universo $\mathcal{X}$?



codifica di Shannon-Fano

$$20 = \mathbb{E}[L^*] \le \mathbb{E}[L^{SF}] \le H_2(X) + 1$$

$$\le \log_2 |\mathcal{X}| + 1$$

$$\log_2 |\mathcal{X}| \geq 20 - 1 = 19 \Rightarrow |\mathcal{X}| \geq 2^{19}$$

(2) Siano $X_1, \dots, X_n$ v.a. independenti e identicamente distribuite su $\mathcal{A} = \{a, b, c, d, e, f\}$, con $p_{X_i} = (1/2, 1/4, 1/8, 1/16, 1/32, 1/32)$

(a) Calcolare $H(X_i)$

(b) Qual è la sequza $(X_1, \dots, X_n)$ più probabile? Quanto vale la sua probabilità?

(c) In relazione all'insieme di tipicità $\mathcal{T}^{(n,\delta)}$ dire se la sequenza di cui al punto appartiene a $\mathcal{T}^{(n,\delta)}$ quando $\delta = 0.1$. Cosa accade se invece $\delta = 1.0$?

(a) $H(X_i) = -\sum_{i=1}^{6} p_i \log p_i = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{16} 4 + \frac{1}{32} \cdot 5 + \frac{1}{32} \cdot 5$

$= 2 - \frac{1}{16} = \underline{1.9375}$ bit

(b) Sequenza $\overbrace{aaaaaa \dots a}^{n \text{ simboli}}$; $p(aa\dots a) = \left(\frac{1}{2}\right)^n = 2^{-n}$

(c) $\mathcal{T}^{(n,\delta)} = \left\{ x \in \mathcal{A}^n : \left| \frac{1}{n} \underbrace{\mathcal{I}(x)}_{-\log p(x)} - H(X) \right| \le \delta \right\}$

$$x \in \mathcal{T}^{(n,\delta)} \iff 2^{-n(H(X)+\delta)} \le p(x) \le 2^{-n(H(X)-\delta)}$$

$H(X) = 1.9375$ bit $\qquad 2^{-n(1)}$

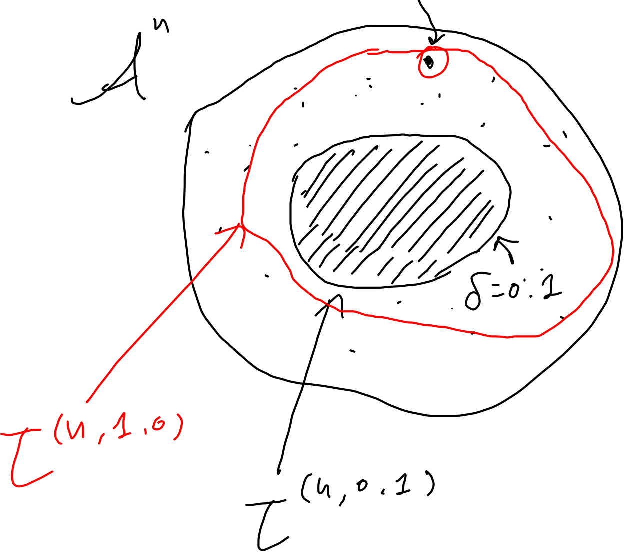
$\delta = 0.1$

$$2^{-n(2.0375)} \overset{\checkmark}{\le} 2^{-n} \overset{\times}{\le} 2^{-n(1.8375)}$$

$x = \overbrace{a \dots a}^{n} \longrightarrow p(x) = 2^{-n}$

NO : $x \notin \mathcal{T}^{(n,\delta)}$

$$1.8375 \overset{\times}{\le} 1 \overset{\checkmark}{\le} 2.0375$$



$$p(\mathcal{T}^{(n,\delta)}) = 1 - \varepsilon_{n,\delta} \xrightarrow[n\to\infty]{} 1$$

Se invece $\delta = 1.0$?

$$2^{-n(2.9375)} \overset{\checkmark}{\le} 2^{-n\cdot 1} \overset{\checkmark}{\le} 2^{-n(0.9375)}$$

OK : $x \in \mathcal{T}^{(n,1.0)}$

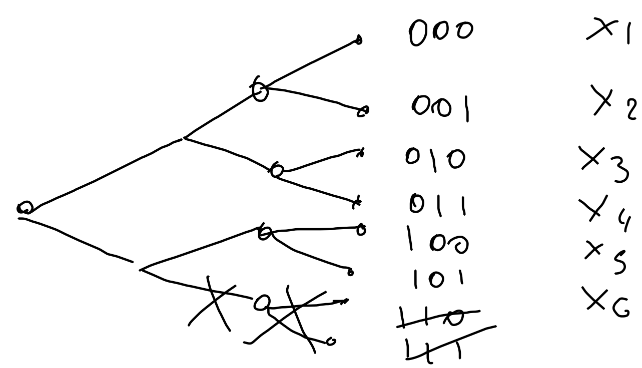
③ Considerare la v.a. X a valori in $\{x_1, x_2, \ldots, x_6\}$ e il codice binario (B-LV)

X	$p(X)$	$Y = \varphi(X)$
x_1	0.5	1
x_2	0.15	000
x_3	0.1	001
x_4	0.1	~~0100~~ $\rightarrow$ 010
x_5	0.1	0110
x_6	0.05	0111

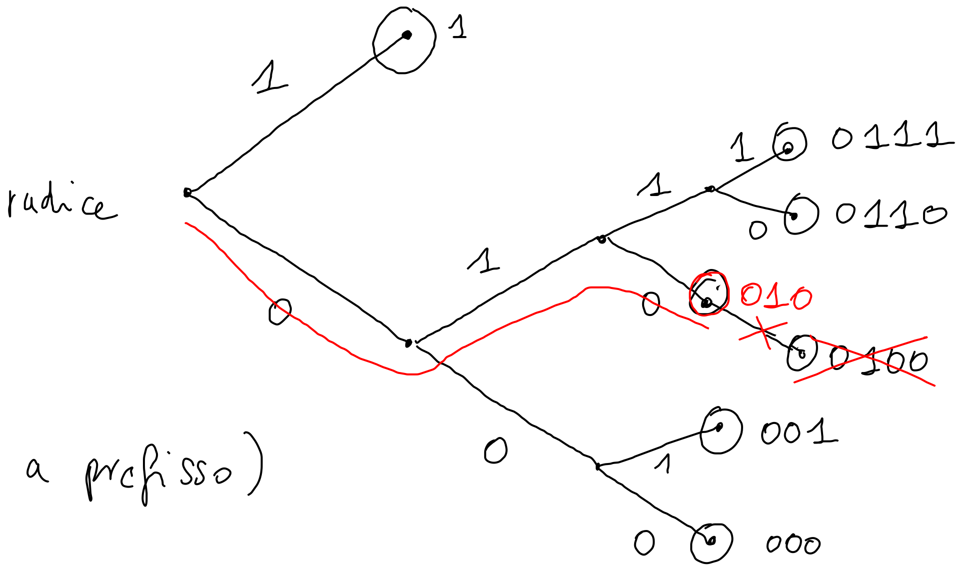
(a) Mostrare che il codice è a prefisso disegnandone una rappresentazione ad albero.

(b) Derivare l'entropia della sorgente e la lunghezza attesa del codice

(c) Determinare se il codice è ottimale (tra quelli u.d.). Se non lo è, fornire un codice con una lunghezza attesa minore.



$\mathbb{E}[L] = 3$



(c'è a prefisso)

$$H(X) = -\sum_{i=1}^{6} p_i \log p_i$$

$$p_i = \Pr[X = x_i]$$

$$= 0.5 \log \frac{1}{0.5} + 0.15 \log \frac{1}{0.15} + \dots$$

$$\approx 2{,}123 \text{ bit}$$

$$\mathbb{E}[L] = \mathbb{E}_X[\varphi(X)] =$$

$$= 0.5 \cdot 1 + 0.15 \cdot 3 + 0.1 \cdot 3 + 0.25 \cdot 4$$

$$= 0.5 + 0.25 \cdot 3 + 0.25 \cdot 4 = \underline{2.25} \geq \underset{\underline{2.123}}{H(X)}$$

($\mathbb{E}[L] \geq H(X)$ per teorema di Shannon)

$$\mathbb{E}[L'] = 0.5 \cdot 1 + 0.35 \cdot 3 + 0.15 \cdot 4 = 2.15 > 2.123$$

④ Dati i tre codici B-LV, per $\mathcal{A} = \{x_0, x_1, \dots, x_5\}$

$\mathcal{B} = \{0, 1\}$

(1)	W	(2) W	(3) W
x_0	11	01	01
x_1	10	10	10
x_2	01	110	001
x_3	001	111	100
x_4	0001	0000	000
x_5	0000	0001	111

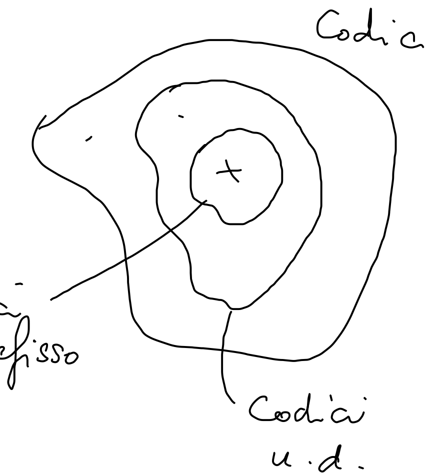
Quali di questi codici sono u.d.?

(1) È a prefisso → è u.d. ✓

(2) È a prefisso → è u.d. ✓

(3) Non è a prefisso. Non so se è u.d. ?

10
100 ← residuo
prefisso comune



Algoritmo di Sardinas-Patterson per decidere se un codice è u.d.

$W = \{01, 10, 001, 100, 000, 111\}$

$R^{(0)} = W$	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$
01	0	1	0
10		01	00
001		00	11
100			⋮
000			
111			

x_3 x_0

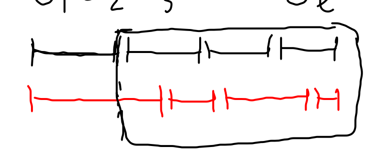
100 | 01 ← ? 10001

10 | 001 ← ?

x_1 x_2

→ Non è u.d.

Se il codice non è u.d. ⤷ $b_1 b_2 b_3 \dots b_\ell$



→ decodifico in 2 modi diversi

Residui di livello 0 : $R^{(0)} = W$

Residui di livello n : $R^{(n)}$

confronto i residui di livello n - 1 con le parole di codice

Si ottiene una qualche parola di codice come residuo (di livello ≥ 1) se e solo se il codice <u>non</u> è u.d.

Es. $W = \{10, 010, 1, 1110\}$

$R^{(0)}$	$R^{(1)}$	$R^{(2)}$
$\boxed{10}$	0	$\boxed{10} \in W$
010	110	(10)
1		
1110		

Mi fermo; il codice non è u.d.

Es. $W = \{0, 001, 101, 11\}$

$R^{(0)}$	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$	$R^{(4)}$
0	01	1	01	1
001			1	01
101				(1)
11				

$R^{(3)} = R^{(4)}$; mi fermo; il codice è u.d.