$$H(X) = -\sum_{i=1}^{K} p_i \log p_i = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

X è una v.a. su un insieme (finito) $\mathcal{X}$ di simboli, con d.d.p. $p = (p_1, p_2, \dots, p_K) \in \mathbb{R}^K$

$(K = |\mathcal{X}|)$

$$H_b(X) = -\sum_i p_i \log_b p_i \qquad \log_b z = \frac{\log_a z}{\log_a b} \Rightarrow H_b(X) = \frac{H_a(X)}{\log_a b}$$

$H(X)$ è una misura dell'incertezza residua (informazione mancante) nella v.a. X

Esempio. $\mathcal{X} = \{1, 2, \dots, 32\}$ e $p = \left(\frac{1}{32}, \frac{1}{32}, \dots, \frac{1}{32}\right)$

Allora $H(X) = -\sum_{i=1}^{32} \frac{1}{32} \log \frac{1}{32} = +\sum_{i=1}^{32} \frac{1}{32} \overbrace{\log 32}^{=5} = 5$ bit

1 : 00000
2 : 00001
. . . .
32 : 11111

Se invece $\mathcal{X} = \{1, \dots, 30\}$?

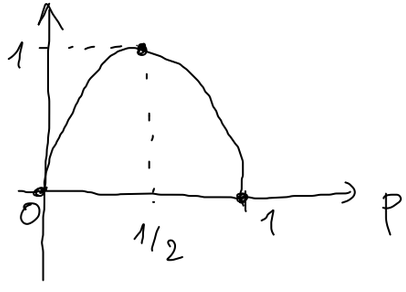
$H(X) \simeq 4.906$

Esercizi. Calcolo dell'entropia. Da entropia a divergenza informazionale. Entropia di funzioni di v.a.

Esempio. $\mathcal{X} = \{0,1\}$ $\quad X = \begin{cases} 0 & \text{con prob. } p \\ 1 & \text{con prob. } 1-p \end{cases}$ (per qualche $p \in [0,1]$)

$$H(X) = -p \log p - (1-p)\log(1-p) \triangleq h_2(p)$$ funzione entropia binaria

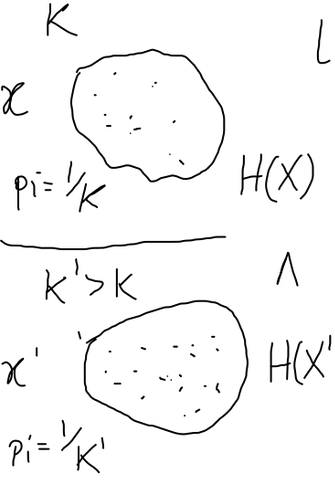
(≥ 0)



Se $p=0$: $-\cancel{0 \log 0} - \cancel{1 \log 1} = 0 = H(X)$

Se $p=1$: $H(X) = 0$

Se $p = 1/2$: $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = \underline{1}$ bit



La formula per l'entropia si può derivare a partire da alcuni assiomi naturali:

→ ① Continuità : $H(X)$ deve essere continua nel vettore $p = (p_1, \ldots, p_K)$

→ ② Monotonia : Se X e X' sono v.a. con distr. uniforme su $\mathcal{X}$, $\mathcal{X}'$ rispettivam. e $|\mathcal{X}| < |\mathcal{X}'|$, allora $H(X) < H(X')$.

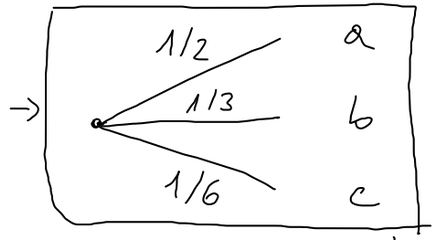
→ ③ Diramazione

$H(X)$ posso scriverla anche come $H(p_X)$

③ Diramazione

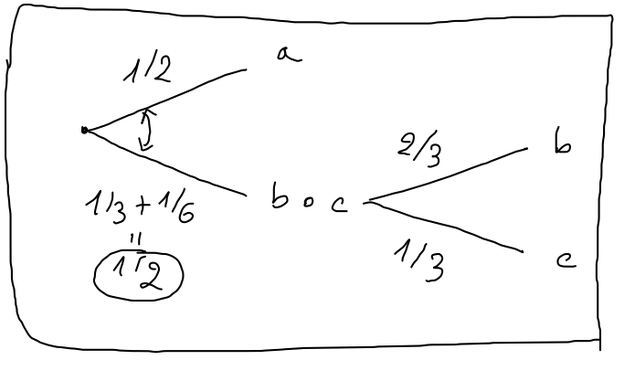
x_i	a	b	c
p_i	1/2	1/3	1/6

$H(X)$



Esempio della proprietà di diramazione

$$H\left(\left(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}\right)\right) = H\left(\left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right) + \frac{1}{2} H\left(\left(\tfrac{2}{3}, \tfrac{1}{3}\right)\right)$$



Si può dimostrare che l'unica funzione (a meno di un fattore costante) H che soddisfa gli assiomi ①, ②, ③ è proprio l'entropia di Shannon $H(X) = -\sum_i p_i \log p_i$

$$D(p\|q) \triangleq \sum_{i=1}^{k} p_i \log p_i/q_i$$

$$I(X;Y) \triangleq D(p_{XY} \| p_X \cdot p_Y) = \sum_{x\in X, y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

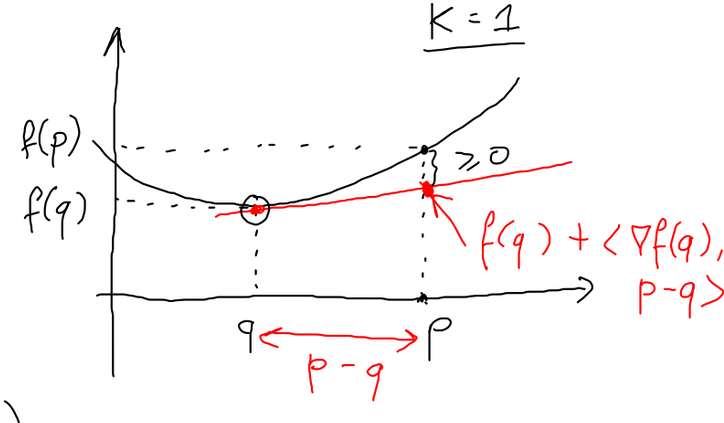
$$H(X) \triangleq I(X;X) = -\sum_i p_i \log p_i$$

$H(p)$

Divergenza di Bregman per una funzione f (derivabile) : $\mathbb{R}^k \to \mathbb{R}$

$$D_f(p\|q) = f(p) - f(q) - \langle \nabla f(q), p-q \rangle$$

$$f(p) \approx f(q) + f'(q)(p-q)$$



Quando f è convessa : $D_f(p\|q) \geq 0$ (per la convessità della f)

$x \log x$

Scelgo:

$$f(x) = \sum_{i=1}^{k} x_i \log x_i$$

Scelgo (entropia negativa) $f(x) = \sum_{i=1}^{k} x_i \log x_i$ (convessa)

$(x \log x)' = 1 + \log x$

$\nabla f(q) = (1+\log q_1, 1+\log q_2, \dots, 1+\log q_k) \in \mathbb{R}^k$

Allora $D_f(p\|q) = f(p) - f(q) - \langle \nabla f(q), p-q \rangle$ (?)

$= \sum_i p_i \log p_i - \sum_i q_i \log q_i - \sum_i (1+\log q_i)(p_i - q_i)$

$= \sum_i p_i \log p_i - \sum_i q_i \log q_i - \underbrace{\sum_i (p_i - q_i)}_{0} - \sum_i (p_i - q_i) \log q_i$

$= \sum_i p_i \log p_i - \sum_i p_i \log q_i = \sum_i p_i \log \frac{p_i}{q_i} = D(p\|q)$.

↑ Divergenza informazionale

Esercizio. Lancio una moneta (onesta) finché non esce testa

Sia X = numero di lanci $\in \{1, 2, 3, \dots\}$

Quanto vale $H(X)$?

(Si usi il fatto $\sum_{n=0}^{\infty} n \cdot r^n = \frac{r}{(1-r)^2}$ per qualunque $r \in [0,1)$)

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$$= -\sum_{k=1}^{\infty} \frac{1}{2^k} \log \frac{1}{2^k} =$$

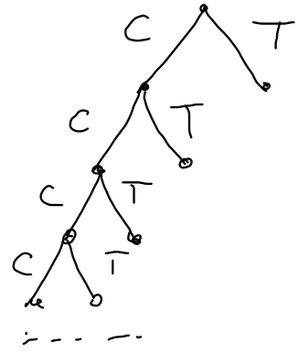
$$= +\sum_{k=1}^{\infty} \frac{1}{2^k} \boxed{\log 2^k} \leftarrow = k \text{ (base 2)}$$

$$= \sum_{k=1}^{\infty} k/2^k = \sum_{k=0}^{\infty} k/2^k$$

$$= \frac{1/2}{(1-1/2)^2} = \frac{1/2}{(1/2)^2} = 2 \text{ bit.}$$

Uso con $r = 1/2$

# lanci (x)	$p(x)$
1	1/2
2	1/4
3	1/8
⋮	⋮
k	$1/2^k$

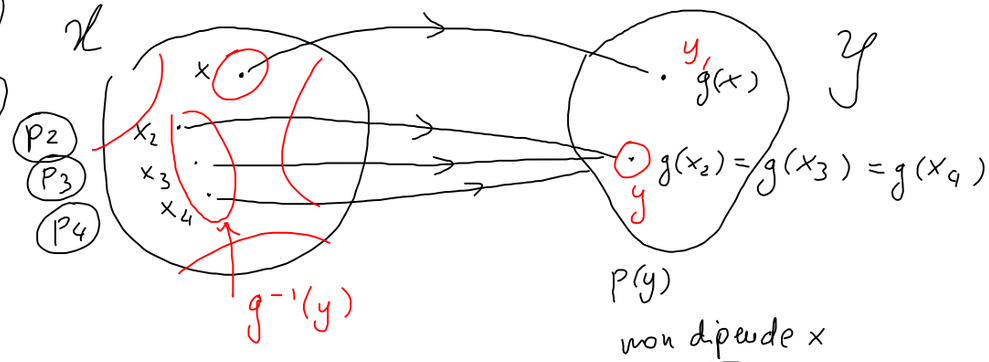


Es. Sia X una v.a. con valori su un sottoinsieme finito di $\mathbb{R}$ ($\mathcal{X}$ o $\mathcal{Y}$)

Che relazione c'è tra $H(X)$ e $H(Y)$ quando:

→ (a) $Y = 2^X$? $H(Y) = H(X)$

→ (b) $Y = \cos(X)$? $H(Y) \leq H(X)$



$y = g(x)$, $g: \mathcal{X} \to \mathcal{Y}$

$\mathcal{X} = \{x_1, \dots, x_K\}$

$$\Pr[Y=y] = p(y) = \sum_{x : g(x)=y} p(x) \quad \Rightarrow \quad \sum_{x:g(x)=y} p(x) \log p(x) \leq \sum_{x:g(x)=y} p(x) \overbrace{\log p(y)}^{\text{non dipende } x} = (\log p(y)) \cdot p(y)$$

$$\Rightarrow \quad H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) = -\sum_{y \in \mathcal{Y}} \sum_{x : g(x)=y} p(x) \log p(x) \geq -\sum_{y \in \mathcal{Y}} p(y) \log p(y) = H(Y)$$

In generale (sia nel caso (a) che nel caso (b)) ho

$H(Y) \leq H(X)$; se g è biettiva tra $\mathcal{X}$ e $\mathcal{Y}$, allora $H(Y) = H(X)$.

① v.a. X = risultato del lancio di un dado a 6 facce $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$

Dado "onesto": $p_X = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ K

$\rightarrow H(X) = -6 \cdot 1/6 \log 1/6 = -\log 1/6 = \log 6 = \log |\mathcal{X}| \approx 2.58$ bit

Dado "truccato": $p_X = (0.001, 0.001, 0.001, 0.001, 0.001, 0.995)$

$\rightarrow H(X) = -\sum_{i=1}^{6} p_i \log p_i \approx \underline{0.057}$ bit

② Qual è il minimo di $H(p_1, p_2, \dots, p_K) = H(p)$

quando p varia nell'insieme di tutte le d.d.p. K-dimensionali?

Determinare i vettori p che determinano il minimo di $H(p)$.

$H(p) = -\sum_{i=1}^{k} p_i \log p_i = \sum_{i=1}^{k} \underbrace{\underbrace{p_i}_{\geq 0} \underbrace{\log (1/p_i)}_{\geq 0}}_{\geq 0} \geq 0$

$H(p)$ è $= 0$ se e solo se $p_i \log (1/p_i) = 0$ per $i = 1, 2, \dots, K$. $\nearrow$ $p_i = 0$

$\rightarrow \log(1/p_i) = 0 \Leftrightarrow p_i = 1$

Quindi p dev'essere della forma:

$$(0, 0, \dots, 0, 1, 0, \dots, 0)$$

Ci sono K vettori di quel tipo:

$(1, 0, \dots, 0)$

$(0, 1, \dots, 0)$ $\Big\}$ $\rightarrow H(p) = 0$

$\dots$

$(0, 0, \dots, 1)$

③ Entropia vs. varianza

Entropia v.a. $X \in \mathcal{X}$ insieme di simboli ; $H(X) = -\sum_{i=1}^{k} p_i \log p_i$

$= \mathbb{E}[-\log p(X)]$

Varianza v.a. $X \in \mathbb{R}$ insieme di numeri reali ; $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$

Se $\mathcal{X} \subseteq \mathbb{R}$, sono applicabili entrambi i concetti. Come si differenziano?

Esempio:

$$X = \begin{cases} 0 & \text{con prob. } 1/2 \\ a\ (\neq 0) & \text{con prob. } 1/2 \end{cases} \qquad \mathcal{X} = \{0, a\} \quad \begin{matrix} 1/2 & 1/2 \end{matrix}$$

$\rightarrow H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = -\log\frac{1}{2} = \log 2 = 1$ bit.

non dipende da a

$\mathrm{Var}(X)$? (a è una costante $\neq 0$)

$$\mathbb{E}[X^2 - 2X\cdot\mathbb{E}X + (\mathbb{E}X)^2] \overset{\text{linearità del valore atteso}}{=} \mathbb{E}[X^2] - 2\,\mathbb{E}[X\cdot\mathbb{E}X] + \mathbb{E}[(\mathbb{E}X)^2]$$

$$X^2 = \begin{cases} 0^2 & \text{con p. } 1/2 \\ a^2 & \text{con p. } 1/2 \end{cases}$$

$$= \mathbb{E}[X^2] - 2\underbrace{[\mathbb{E}X]\cdot\mathbb{E}[X]}_{(\mathbb{E}X)^2} + (\mathbb{E}X)^2 = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{1}{2}a^2 - \left(\frac{a}{2}\right)^2 = \frac{a^2}{4}.$$

dipende da a

X, Y v.a.

v.a. indipendenti : $p_{XY} = p_X \cdot p_Y$ $\quad p(x,y) = p(x)p(y) \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}$

$\Leftrightarrow D(p_{XY} \| p_X \cdot p_Y) = 0 \Leftrightarrow I(X;Y) = 0$

vs ⇓ ✳

v.a. scorrelate : $\mathrm{Cov}(X,Y) = 0$

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

$$(\mathrm{Var}(X) = \mathrm{Cov}(X,X))$$

Esempio in cui X e Y sono scorrelate ma non indipendenti.

p_{XY}	-1	0	$+1$	p_X
-1	0	$1/4$	0	$1/4$
$X \to 0$	$1/4$	0	$1/4$	$1/2$
$+1$	0	$1/4$	0	$1/4$
$p_Y \to$	$(1/4$	$1/2$	$1/4)$	

(Y: -1, 0, $+1$)

X e Y non sono indipendenti; per es.

$$p(x=-1, y=-1) \neq p(x=-1) \cdot p(y=-1)$$

$$0 \qquad\qquad 1/4 \qquad\qquad 1/4$$

$$I(X;Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

$$= \cancel{1/4} \cdot \log \frac{\cancel{1/4}}{1/2 \cdot \cancel{1/4}} \cdot \cancel{4} = \log 2 = 1$$

$$\mathbb{E}X = \tfrac{1}{4}(-1) + \cancel{\tfrac{1}{2}\,0} + \tfrac{1}{4}(+1) = 0$$

$$\mathbb{E}Y = 0$$

$$\Rightarrow \mathrm{Cov}(X,Y) = \mathbb{E}[XY] = \tfrac{1}{4}\cdot 0 + \tfrac{1}{4}\cdot 0 + \dots + \tfrac{1}{4}\cdot 0 = 0$$

$\Rightarrow$ X e Y sono scorrelate.

④ X, Y v.a.

p_{XY}	c	d	p_X
a	0	1/8	1/8
b	3/4	1/8	7/8
p_Y	3/4	1/4	

(X: a, b; Y: c, d)

$p_X = (1/8, 7/8)$

$$H(X) = -\tfrac{1}{8}\log\tfrac{1}{8} - \tfrac{7}{8}\log\tfrac{7}{8} \approx 0.544 \text{ bit}$$

Calcolare $H(X)$, $H(X|Y)$, e $I(X;Y)$.

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p(y)\, H(X|Y=y) =$$

$$= 3/4 \cdot H(X|Y=c) + 1/4\, H(X|Y=d) =$$

$$p_{X|Y=c} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad p_{X|Y=d} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

$\log 2 = 1$ bit

$$= \tfrac{3}{4}\underbrace{\left[H((0,1))\right]}_{0} + \tfrac{1}{4}\underbrace{\left[H((1/2,1/2))\right]}_{1/4\,\cdot\,1} = \frac{1}{4}$$

$$I(X;Y) = H(X) - H(X|Y) \cong 0.544 - 1/4 .$$

$H(Y) = \log 4 = 2$ bit

⑤ V.a. X, Y

Calcolare $H(X)$, $H(X|Y)$, e $I(X;Y)$.

p_{XY} (X \ Y)	1	2	3	4	p_X
1	1/8	1/16	1/16	1/4	1/2
2	1/16	1/8	1/16	0	1/4
3	1/32	1/32	1/16	0	1/8
4	1/32	1/32	1/16	0	1/8
p_Y	1/4	1/4	1/4	1/4	

$$H(X) = -\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{2}{8}\log\tfrac{1}{8} = \tfrac{1}{2} + \tfrac{1}{4}\cdot 2 + \tfrac{2}{8}\cdot 3 = \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{3}{4} = \tfrac{7}{4} .$$

$$H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y=y) = \tfrac{1}{4} H(X|Y=1) + \tfrac{1}{4} H(X|Y=2) + \tfrac{1}{4} H(X|Y=3) + \tfrac{1}{4} H(X|Y=4)$$

$$= 11/8 .$$

$$I(X;Y) = H(X) - H(X|Y) = \tfrac{7}{4} - \tfrac{11}{8} = \tfrac{14}{8} - \tfrac{11}{8} = \tfrac{3}{8} .$$

$(X \oplus Y) \oplus Y = X$

È possibile definire tre v.a. X, Z, Y :

$\begin{cases} I(X;Z|Y) > 0 & \checkmark \\ I(X;Z) = 0 & \checkmark \end{cases}$

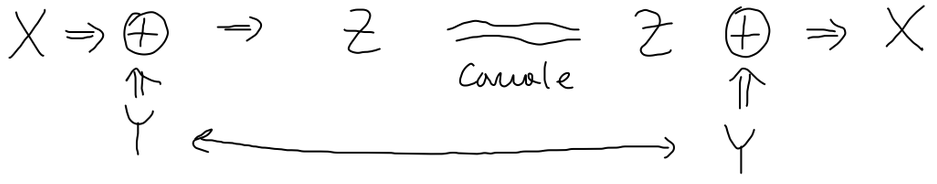
$Z = X \oplus Y$

(6) SIGSALY

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

$p_X = (\alpha, 1-\alpha)$

per qualche $\alpha \in [0,1]$



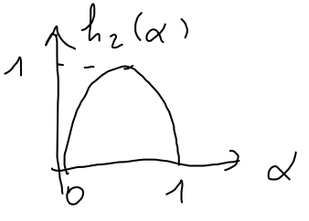
$p_Y = (1/2, 1/2)$

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

(a) Supp. che $p_X = (\alpha, 1-\alpha)$, $p_Y = (1/2, 1/2)$ e che X e Y siano v.a. indipendenti

Dimostrare che $H(X|Z) = H(X)$ ("Z non fornisce alcuna informazione su X")

(b) Supp. che $p_Y = (1/4, 3/4)$. Cosa cambia?



(a) $H(X) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) = h_2(\alpha)$

$H(X) - H(X|Z) = I(X;Z)$ → dipende da p_{XZ}

$I(X;Z) = 0$

$p_{XZ} = p_X \cdot p_Z$

X

p_{XY}	0	1	p_X
0	$\alpha/2$	$\alpha/2$	α
1	$(1-\alpha)/2$	$(1-\alpha)/2$	$1-\alpha$
p_Y	$1/2$	$1/2$	

($Z=0$: $\alpha/2$, $(1-\alpha)/2$; $Z=1$: $\alpha/2$, $(1-\alpha)/2$)

X

p_{XZ}	0	1	p_X
0	$\alpha/2$	$\alpha/2$	α
1	$(1-\alpha)/2$	$(1-\alpha)/2$	$1-\alpha$
p_Z	$1/2$	$1/2$	

$$I(X;Z|Y) = \tfrac{1}{2}\, I(X;Z|Y=0) + \tfrac{1}{2}\, I(X;Z|Y=1)$$

$$= \tfrac{1}{2}\, \overbrace{I(X;X)}^{\geq 0} + \tfrac{1}{2}\, \overbrace{I(X;1-X)}^{\geq 0}$$

$$\geq \tfrac{1}{2}\, H(X)$$

Se $Y=0$, $X=Z$

Se $Y=1$, $X=1-Z$

Quindi per esempio con $\alpha = 1/2$,

ho $I(X;Z|Y) \geq \tfrac{1}{2} H(X) = \tfrac{1}{2} > 0$.

(10) SIGSALY

$X \oplus \underbrace{Y \oplus Y}_{Z} = X$

$X \to \oplus \to Z \rightsquigarrow \approx \to Z \to \oplus \to X$, with Y entering the first $\oplus$ and Y entering the second $\oplus$.

(a) $p_X = (\underbrace{\alpha}_{X=0}, \underbrace{1-\alpha}_{X=1})$

$p_Y = (1/2, 1/2)$

$\Rightarrow$ $I(X;Z) = 0$

$I(X;Z|Y) > 0$

(b) $p_X = (\alpha, 1-\alpha)$

$p_Y = (1/4, 3/4)$

$\Rightarrow$ Cosa cambia?

X e Z sono v.a. indipendenti? NO $\nearrow$ $I(X;Z) > 0$

$p_X \cdot p_Y =$

X \ Y	0	1
0	$\alpha/4$	$3\alpha/4$
1	$(1-\alpha)/4$	$3(1-\alpha)/4$

$Z=1$, $Z=0$

p_{XZ}

X \ Z	0	1
0	$\alpha/4$	$3\alpha/4$
1	$3(1-\alpha)/4$	$(1-\alpha)/4$

NON può essere

$p_{XZ} \overset{?}{=} p_X \cdot p_Z$

$\begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \begin{array}{|c} (1/4 \quad 3/4) \\ \hline \alpha/4 \quad 3\alpha/4 \\ \hline (1-\alpha)/4 \\ \hline \text{contraddizione} \end{array}$

Esercizi. Mutua informazione di 3 v.a. e applicazioni. Teorema della segretezza perfetta di Shannon.

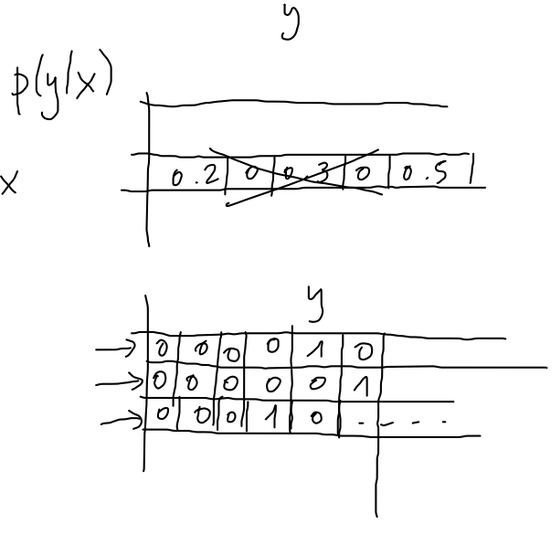
① Mostrare $H(Y|X)=0 \iff$ per ogni $x \in \mathcal{X}$ con $p(x)>0$

esiste uno e un solo y tale che $p(y|x)=1$

(in altre parole, Y è funzione di X).

(Generalizza il fatto che $H(Y)=0 \iff Y$ è una costante)

Dim. $H(Y|X)=0 \iff \sum_{x\in\mathcal{X}} \underbrace{p(x)}_{\geq 0} \underbrace{H(Y|X=x)}_{\geq 0} = 0$ $\quad \left(\mathbb{E}_x[H(Y|X=x)]\right)$



$\iff \sum_{x\in\mathcal{X}:p(x)>0} \overbrace{p(x)}^{\geq 0} \overbrace{H(Y|X=x)}^{\geq 0} = 0$

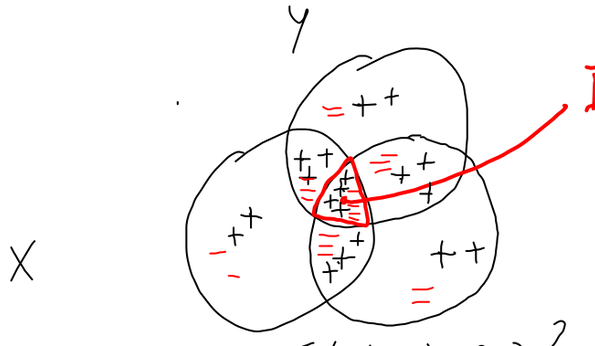
$\iff \forall x: p(x)>0$ ho $H(Y|X=x)=0$

$\iff \forall x: p(x)>0$ ho $-\sum_{y\in\mathcal{Y}} \underbrace{p(y|x)}_{\geq 0} \underbrace{\log p(y|x)}_{\leq 0} = 0$

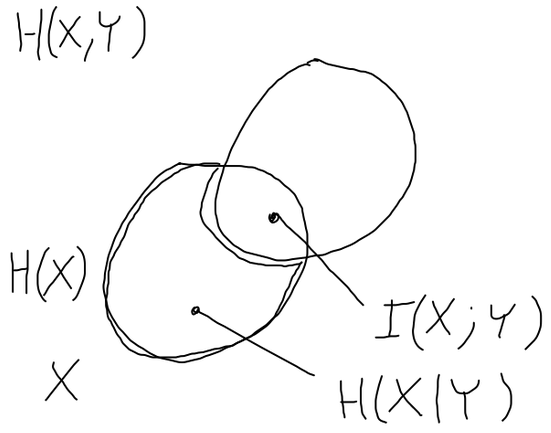
$\iff \forall x: p(x)>0$ $\forall y \in \mathcal{Y}$ ho $p(y|x)=0$ oppure $p(y|x)=1$

$\iff Y$ è funzione di X.

② "Mutua informazione" tra 3 v.a.



$I(X;Y;Z)$ $(\mu(X \cap Y \cap Z))$



Come definisco $I(X;Y;Z)$?

Una possibilità è la seguente:

$$I(X;Y;Z) \overset{def}{=} H(X,Y,Z) - H(X,Y) - H(X,Z) - H(Y,Z) + H(X) + H(Y) + H(Z)$$

Attenzione: $I(X;Y;Z)$ NON è una divergenza informazionale

$$\mu(X \cap Y \cap Z) + \mu(X \cup Y) + \mu(X \cup Z) + \mu(Y \cup Z) \overset{\checkmark}{=} \mu(X \cup Y \cup Z) + \mu(X) + \mu(Y) + \mu(Z)$$

$$\left(I(X;Y) \overset{def}{=} D(p_{XY} \| p_X \cdot p_Y) \geq 0 \right)$$

$;$ $\rightarrow$ $\cap$

$|$ $\rightarrow$ $\setminus$

$,$ $\rightarrow$ $\cup$

μ

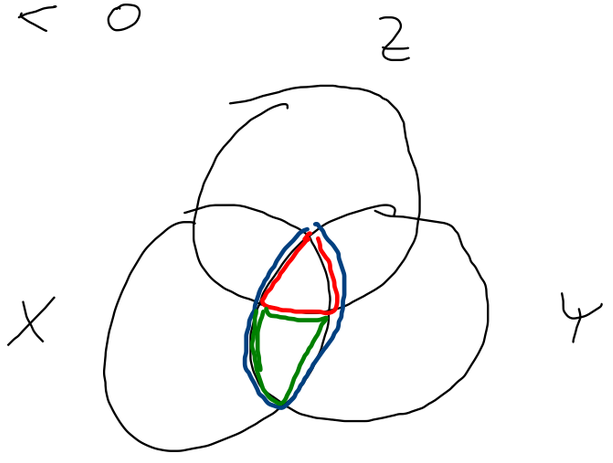
$\mu(S_X) = H(X) \geq 0$

$\mu(S_X \cap S_Y) = I(X;Y) \geq 0$

$$\rightarrow \quad I(X;Y;Z) = I(X;Y) - I(X;Y|Z) < 0$$

$I(X;Y) = 0$ ↑ $I(X;Y|Z) > 0$ ↗

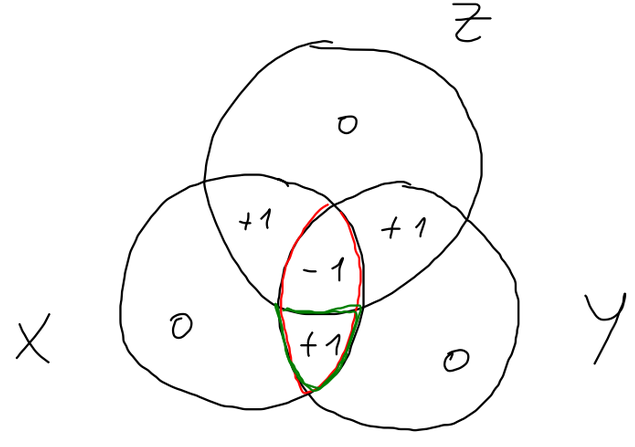
si può avere simultaneamente



Esempio : X uniforme su $\{0,1\}$

Z uniforme su $\{0,1\}$

$Y = X \oplus Z$



$H(X) = 1$

$I(X;Y) = 0$ $\quad I(X;Y|Z) = 1$

$I(X;Y;Z) = -1$

③ Esempio di 3 v.a. X, Y, Z con $I(X;Y) > 0$ e $I(X;Y|Z) = 0$.

Consideriamo 3 v.a. X, Y, Z in catena di Markov $X \to Z \to Y$

con X e Y non indipendenti

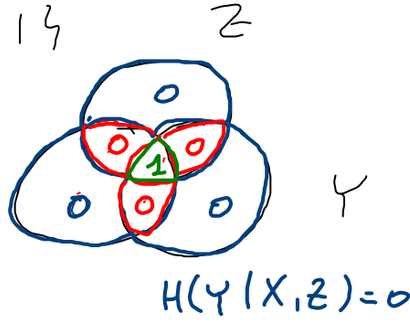
Allora, X e Y non indipendenti $\to$ $I(X;Y) > 0$

D'altra parte per Markovianità, $I(X;Y|Z) = 0$

→ Scenario concreto: Prendo X con distribuzione uniforme su $\{0,1\}$

Prendo $Z = X$

Prendo $Y = Z = X$



$I(X;Y) = I(X;X) = H(X) = 1 > 0$

$H(Z) = H(X) = H(Y) = 1$

$I(X;Y|Z) = \underbrace{H(X|Z)}_{0} - \underbrace{H(X|Y,Z)}_{0} = 0$ → Vale la catena $X \to Z \to Y$.

→ $I(X;Y;Z) = 1$

④ Teorema della segretezza perfetta (Shannon 1949)

X, Y, Z v.a.

X : testo in chiaro

Y : testo cifrato

Z : chiave di cifratura

Uno schema di cifratura "ideale" dovrebbe avere queste proprietà:

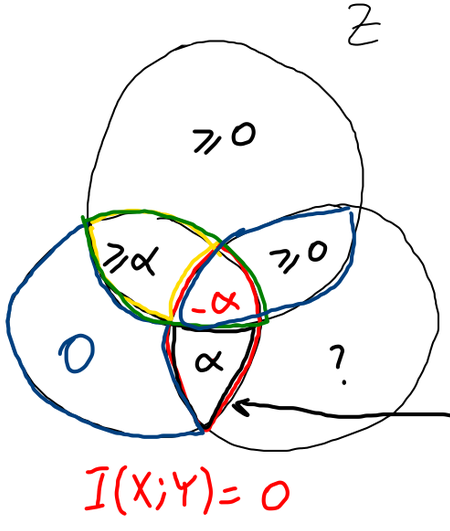
(1) Decifrabilità : $H(X|Y,Z) = 0$ (X è funzione della coppia (Y,Z))

(2) Segretezza perfetta : $I(X;Y) = 0$

Teorema : Per qualunque schema di cifratura con le proprietà (1) e (2), si ha:

$H(Z) \geq H(X)$. (in pratica: la chiave deve essere lunga quanto il testo in chiaro)

Dim.



$I(X;Z) \geq 0$ $\quad I(Y;Z) \geq 0$

$H(X) = 0 + \alpha - \alpha + \bullet$ ($\bullet \geq \alpha$)

$H(Z) = \geq 0 + \bullet + \geq 0$

$I(X;Y|Z) \geq 0$

$\longrightarrow H(Z) \geq H(X)$

(Nota. Nel one-time pad, si ha $H(Z) = H(X)$)

QED.

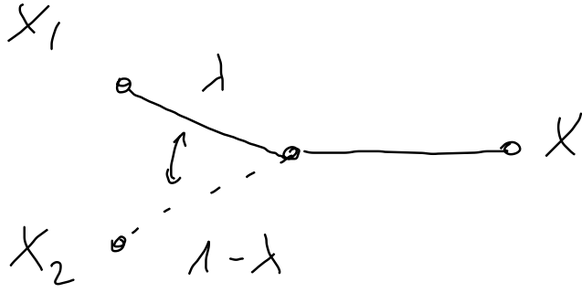
Sia $\lambda \in (0,1)$

(5) Consideriamo due v.a. X_1 e X_2

Definiamo una terza v.a. X, come $X = \begin{cases} X_1 & \text{con prob. } \lambda \\ X_2 & \text{con prob. } 1-\lambda \end{cases}$

Dimostrare che

$$H(X) \geq \lambda H(X_1) + (1-\lambda) H(X_2)$$



Teoria dell'informazione e inferenza statistica

Sia $\{p_\vartheta(x)\}_{\vartheta}$ una famiglia di d.d.p. parametrizzata da un parametro $\vartheta \in \mathbb{R}^k$

Esempio: Lancio di una moneta

$$p_\vartheta(x) = \begin{cases} \vartheta & \text{se } x \text{ è Testa } (0) \\ 1-\vartheta & \text{se } x \text{ è croce } (1) \end{cases}$$

Sia X una v.a. campionata da p_ϑ : $X \sim p_\vartheta$ $\quad X \sim p_\vartheta^n$

Una statistica è una funzione $T(X)$ del campione X

Esempio: media campione : $T(\underbrace{X_1, X_2, \dots, X_n}_{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$

Poiché $T(X)$ è una funzione del campione, vale la catena di Markov : $\vartheta \to X \to T(X)$

$$\left(Pr[T(X) \mid \vartheta, X] = Pr[T(X) \mid X] \right)$$

$\Rightarrow$ Per il 2° teorema di elaborazione dati, $I(\vartheta; T(X)) \leq I(\vartheta; X)$

$\overset{?}{=}$

Esercizi. Statistiche sufficienti, massima verosimiglianza. Esempi di codifica di sorgente.

Quando $I(\vartheta; T(X)) = I(\vartheta; X)$, $T(X)$ è detta statistica sufficiente

Questo è vero quando

$$Pr[X \mid \vartheta, T(X)] = Pr[X \mid T(X)]$$

ovvero quando vale la catena di Markov: $\vartheta \to T(X) \to X$

Esempio. $X_1, X_2, \dots, X_n$ v.a. indipendenti, binarie, identicamente distribuite; con $p(1) = \vartheta$, $p(0) = 1-\vartheta$ $(\vartheta \in [0,1])$

$$Pr[(X_1, \dots, X_n) = (x_1, \dots, x_n)] = \vartheta^{\sum_i x_i} (1-\vartheta)^{n - \sum_i x_i}$$

$$T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$$

$$\underbrace{Pr\left[(X_1, \dots, X_n) = (x_1, \dots, x_n) \,\middle|\, \sum_{i=1}^n X_i = k\right]}_{Pr[X \mid T(X)]} = \begin{cases} 1/\binom{n}{k} & \text{se } \sum_{i=1}^n x_i = k \\ 0 & \text{se } \sum_{i=1}^n x_i \neq k \end{cases}$$

non dipende da ϑ

$\Rightarrow \sum_i X_i$ è una statistica sufficiente

$$Pr[(X_1, X_2, \dots, X_5) = 1001] = \vartheta(1-\vartheta)^3\vartheta = \vartheta^2(1-\vartheta)^3$$

$$Pr[\underbrace{0100}_{\binom{5}{2}}1] = \vartheta^2(1-\vartheta)^3$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Stima a massima verosimiglianza (MLE: Maximum Likelihood Estimation)

Scenario: Ho una famiglia di d.d.p $\{p_\vartheta(x)\}_\vartheta$; esiste un modello corretto $p_{\vartheta^*}(x)$ ma non conosco ϑ^*

Qual è il valore di ϑ maggior consistente con le osservazioni?

Una possibilità è di misurare la consistenza di ϑ (rispetto al valore corretto ϑ^*) come segue:

$$D(p_{\vartheta^*} \| p_\vartheta) = \mathop{\mathbb{E}}_{X\sim p_{\vartheta^*}}\left[\log \frac{p_{\vartheta^*}(X)}{p_\vartheta(X)}\right] = \underbrace{\mathop{\mathbb{E}}_{X\sim p_{\vartheta^*}}[\log p_{\vartheta^*}(X)]}_{\text{non dipende da } \vartheta} - \mathop{\mathbb{E}}_{X\sim p_{\vartheta^*}}[\log p_\vartheta(X)]$$

$= \text{costante}$ (rispetto a ϑ) $- \mathop{\mathbb{E}}_{X\sim p_{\vartheta^*}}[\log p_\vartheta(X)]$

Non la conosco, ma posso stimarla:

per la legge dei grandi numeri,

$$\frac{1}{n}\sum_{i=1}^n \log p_\vartheta(X_i) \xrightarrow{\text{prob.}} \mathop{\mathbb{E}}_{X\sim p_{\vartheta^*}}[\log p_\vartheta(X)] \quad (n\to\infty)$$

Considero:

$$\hat{D}(p_{\vartheta^*} \| p_\vartheta) = \text{costante (rispetto a } \vartheta) - \frac{1}{n}\sum_{i=1}^n \log p_\vartheta(X_i)$$

Scelgo ϑ in modo da minimizzare $\hat{D}(p_{\vartheta^*} \| p_\vartheta)$

In generale,

$$\min_\vartheta \hat{D}(p_{\vartheta^*} \| p_\vartheta) \iff \min_\vartheta \left[-\frac{1}{n} \sum_{i=1}^n \log p_\vartheta(X_i) \right]$$

$$\iff \min_\vartheta \left[-\frac{1}{n} \log \prod_{i=1}^n p_\vartheta(X_i) \right]$$

$$\iff \max_\vartheta \left[\log \prod_{i=1}^n p_\vartheta(X_i) \right]$$

$\log(\cdot)$ è una funz. monotona

$$\iff \boxed{\max_\vartheta \prod_{i=1}^n p_\vartheta(X_i)}$$

→ funzione di verosimiglianza (likelihood)

$$\max_\vartheta \mathcal{L}(\vartheta | X)$$

Esempio: $p_\vartheta(X) = \begin{cases} \vartheta & \text{se } X=1 \\ 1-\vartheta & \text{se } X=0 \end{cases}$ $(\vartheta \in [0,1])$

$$\mathcal{L}(\vartheta | \overbrace{X_1, \dots, X_n}^{X}) = \prod_{i=1}^n p_\vartheta(X_i) = \vartheta^k (1-\vartheta)^{n-k}$$ dove $k = \sum_{i=1}^n X_i$

Osservazione: $\mathcal{L}(\vartheta | X) \geq 0$ sempre

$\mathcal{L}(0|X) = 0 = \mathcal{L}(1|X)$

Calcolo $\mathcal{L}'(\vartheta | X) = k\vartheta^{k-1}(1-\vartheta)^{n-k} - (n-k)\vartheta^k (1-\vartheta)^{n-k-1}$

$$\mathcal{L}'(\vartheta|X) = k\vartheta^{k-1}(1-\vartheta)^{n-k} + (n-k)\vartheta^{k}(1-\vartheta)^{n-k-1}$$

$$= \underbrace{\vartheta^{k-1}}\ \underbrace{(1-\vartheta)^{n-k-1}}\ \underbrace{\left[k(1-\vartheta) - (n-k)\vartheta\right]}$$

Quando ho $\mathcal{L}'(\vartheta|X) = 0$? $\vartheta = 0$ oppure $\vartheta = 1$

oppure $k(1-\vartheta) - (n-k)\vartheta = 0$

$\Updownarrow$

$$k - k\vartheta - (n-k)\vartheta = 0$$

$$k - n\vartheta = 0 \quad \Rightarrow \quad \boxed{\vartheta = k/n}$$

Esercizio. Codifica run-length

Siano $X_1, \dots, X_n$ n v.a. $\in \{0,1\}$, non necessariamente indipendenti

Sia $R = (R_1, R_2, \dots)$ la sequenza delle run

$X =$ 0 0 0 1 1 0 0 1 0 0 ↕ codifica

1 1 1 0 0 1 1 0 1 (1) $\rightarrow R = (3, 2, 2, 1, 2)$

3 2 2 1 2

Si confrontino:

→ $H(X_1, \dots, X_n)$ (cioè $H(X)$), $H(R_1, R_2, \dots)$ (cioè $H(R)$), e $H(X_n, R)$

e limitare le differenze tra queste 3 quantità.

Osservo che $R = (R_1, R_2, \dots)$ è una funzione di X; quindi $H(R) \leq H(X)$

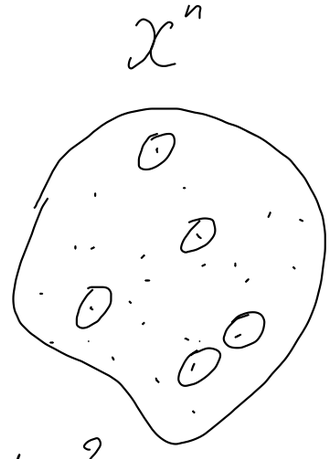
Inoltre: X è una funzione della coppia (X_n, R) $\rightarrow H(X) \leq H(X_n, R)$

(X_n, R) è una funzione di X $\rightarrow H(X_n, R) \leq H(X)$

$\Big\} H(X_n, R) = H(X)$

$$H(R) \leq H(X) = H(X_n, R) = H(R) + H(X_n | R) \leq H(R) + \underbrace{H(X_n)}_{\leq 1 = \log_2 |\{0,1\}|} = H(R) + 1 \dots$$

Esercizio. Siano $X_1, X_2, \dots, X_n$ di v.a. indipendenti e identicamente distribuite, con entropia $\bar{H}$.



Sia $C_n(t) = \{ (x_1, x_2, \dots, x_n) \in \mathcal{X}^n : p(x_1, \dots, x_n) \geq \underline{2^{-nt}} \}$

(a) Mostrare che $|C_n(t)| \leq 2^{nt}$

(b) Per quali valori di t si ha $\Pr[(X_1, \dots, X_n) \in C_n(t)] \xrightarrow{n\to\infty} 1$?

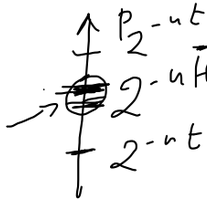
(a) $1 \geq \Pr[(X_1, \dots, X_n) \in C_n(t)] \geq |C_n(t)| \cdot \min_{x \in C_n(t)} p(x) \geq |C_n(t)| \cdot 2^{-nt}$

$\Rightarrow |C_n(t)| \leq 2^{nt}$.

(b) Per il principio di equipartizione asintotica:

$\underbrace{-\frac{1}{n} \log p(X)}_{\text{autoinformaz. normalizzata}} \xrightarrow{\text{prob.}} \bar{H}$ $\Rightarrow$ $p(X)$ si comporta come $2^{-n\bar{H}}$

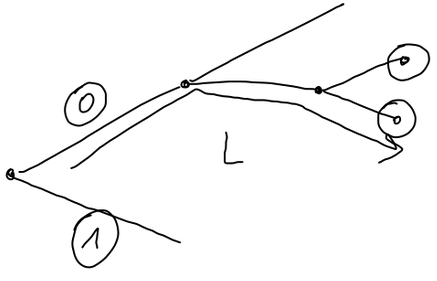
Quindi $\Pr[p(X) \geq 2^{-nt}] \to 1$ se $t > \bar{H}$, $\to 0$ altrimenti (se $t < \bar{H}$)



(1) Gioco "Twenty Questions"

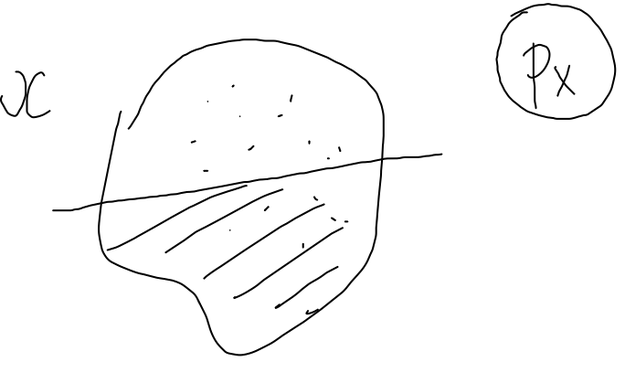
Gioco: giocatore A sceglie un elemento X da un insieme universo $\mathcal{X}$

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

$$20 = \mathbb{E}[L^*] \leq \mathbb{E}[L^{SF}] \leq H_2(X) + 1 \leq \log_2 |\mathcal{X}| + 1$$

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

(2) Siano $X_1, \dots, X_n$ v.a. indipendenti 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 sequenza $(X_1, \dots, X_n)$ più probabile? Quanto vale la sua probabilità?

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

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

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

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

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

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

$H(X) = 1.9375$ bit

$\delta = 0.1$

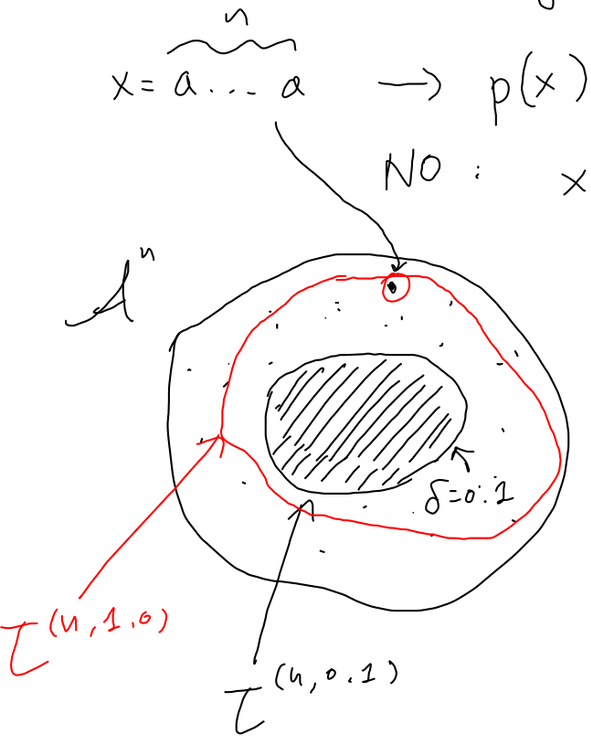
$x = \underbrace{a \dots a}_{n} \longrightarrow p(x) = 2^{-n}$

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

$2^{-n(1)}$

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

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



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

Se invece $\delta = 1.0$?

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

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

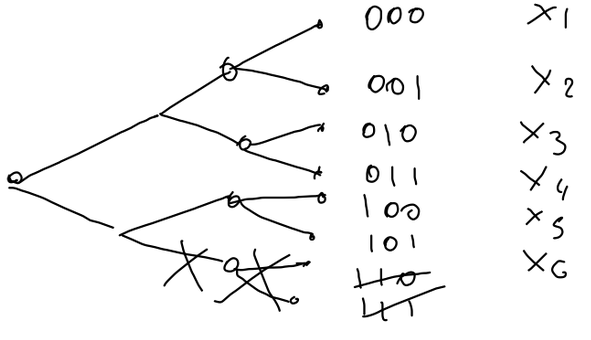
③ Considerare la v.a. X a valori in $\{x_1, x_2, \dots, x_6\}$ e il codice binario (B-LV)

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~~ → 010
x_5	0.1	0110
x_6	0.05	0111

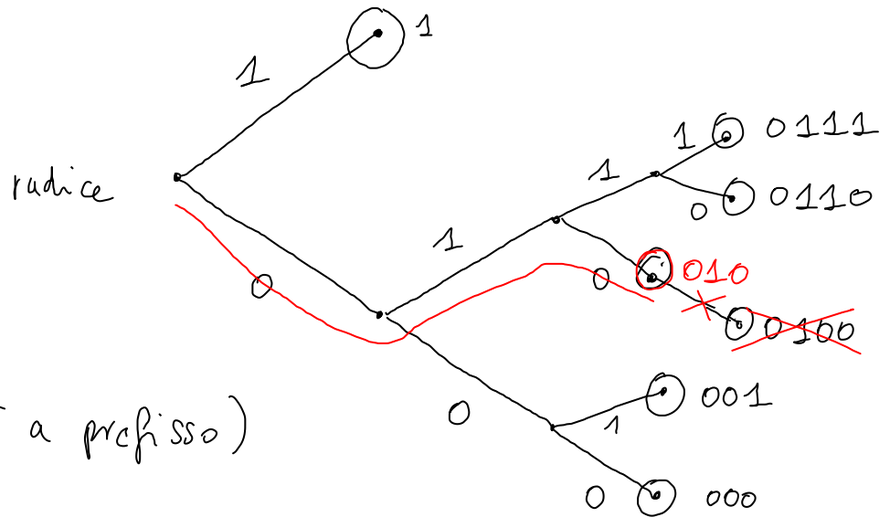
(a) Mostrare che il codice è a prefisso disegnandone una rappresentazione ad albero.

(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.



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



(è a prefisso)

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

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

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

$$\approx 2.123 \quad \text{bit}$$

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

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

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

($\mathbb{E}[L] \geq H(X)$ 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 $A = \{x_0, x_1, \dots, x_5\}$

$B = \{0, 1\}$

Quali di questi codici sono u.d.?

(1)

	W
x_0	11
x_1	10
x_2	01
x_3	001
x_4	0001
x_5	0000

È a prefisso → è u.d. ✓

(2)

W
01
10
110
111
0000
0001

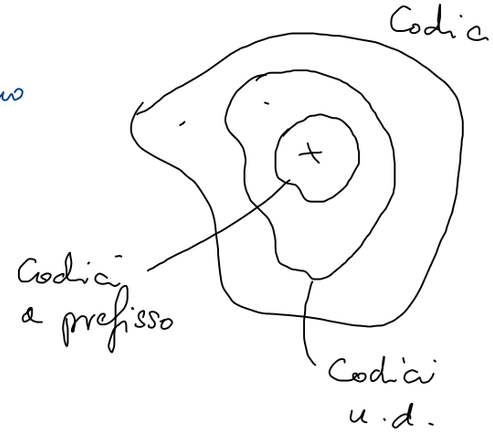
È a prefisso → è u.d. ✓

(3)

W
01
10
001
100
000
111

Non è a prefisso
Non so se è u.d. ?

10 — residuo
100 — prefisso comune

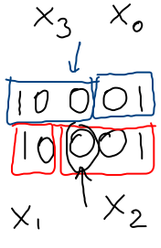


Algoritmo di Sardinas-Patterson per decidere se un codice è u.d.

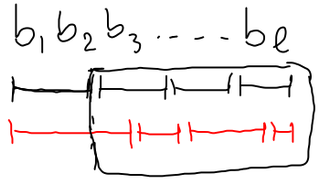
$W = \{01, 10, 001, 100, 000, 111\}$

$R^{(0)} = W$

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



Se il codice non è u.d. ↝ 10001 → Non è u.d.



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 non è u.d.

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

$R^{(0)}$ | $R^{(1)}$ | $R^{(2)}$

10 | 0 | 10 $\in W$

010 | 110 | (10)

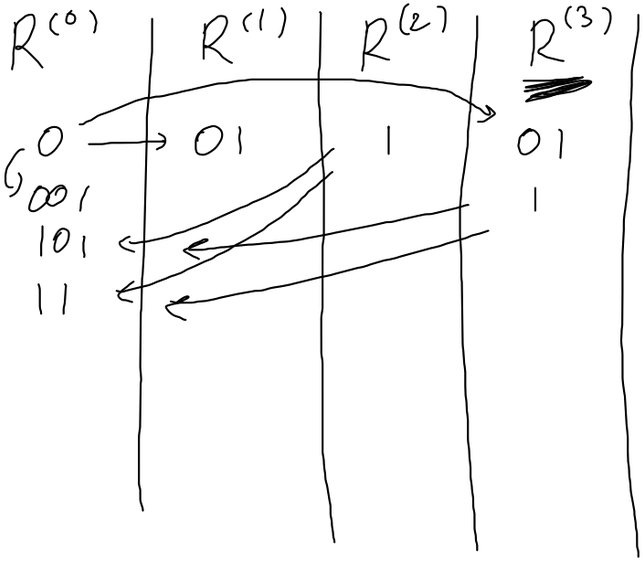
1

1110

Mi fermo;

il codice non è u.d.

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



$R^{(3)} = R^{(4)}$;

mi fermo ; il codice è u.d.

① Siano X, Y due v.a. sullo stesso alfabeto $A = \{a_1, a_2, a_3, a_4, a_5\}$ con d.d.p. p_X, p_Y rispettivamente.

Consideriamo due funzioni di codifica φ_1, φ_2 per gli elementi di A

simbolo(x)	$p_X(x)$	$p_Y(x)$	$\varphi_1(x)$	L	$\varphi_2(x)$	
a_1	1/2	1/2	0	(1)	0	(1)
a_2	1/4	1/8	10	(2)	100	(3)
a_3	1/8	1/8	110	(3)	101	(3)
a_4	1/16	1/8	1110	(4)	110	(3)
a_5	1/16	1/8	1111	(4)	111	(3)

(1) Calcolare $H(X)$, $H(Y)$, $D(p_X \| p_Y)$, $D(p_Y \| p_X)$

(2) Mostrare che la lunghezza attesa delle parole di codice di φ_1 utilizzata per la v.a. X coincide con $H(X)$ e che quindi la codifica φ_1 è ottima per X. Mostrare che φ_2 è ottima per Y.

(3) Si supponga di utilizzare la codifica φ_2 per la v.a. X. Che lunghezza attesa delle parole di codice otteniamo? Di quanto è maggiore di $H(X)$?

Esercizi. Esempi su Fano, Huffman, Ziv-Lempel.

Si evidenzi la relazione con $D(p_X \| p_Y)$.

(1) $p_X = (1/2, 1/4, 1/8, 1/16, 1/16)$, $p_Y = (1/2, 1/8, 1/8, 1/8, 1/8)$

$H(X) = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8\, 3 + 2/16 \cdot 4 = 1/2 + 1/2 + 3/8 + 1/2 = 15/8 = 2 - 1/8$. (bit)

$H(Y) = 1/2 \cdot 1 + 4/8 \cdot 3 = 1/2 + 12/8 = 1/2 + 3/2 = 2$ bit.

$D(p_X \| p_Y) = \sum_{i=1}^{K} (p_X)_i \log \frac{(p_X)_i}{(p_Y)_i} = \cancel{1/2 \cdot \log 1} + 1/4 \log 2 + \cancel{1/8 \log 1} + 2/16 \log 1/2$

$= 1/4 + 1/8 (-1) = 1/8$.

(φ_2)

$D(p_Y \| p_X) = \cancel{1/2 \log 1} + 1/8 \log 1/2 + \cancel{1/8 \log 1} + 2/8 \log 2 = -1/8 + 1/4 = 1/8$

(2) (φ_1)

$\mathbb{E}[L^{(1)}] = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/16 \cdot 4 + 1/16 \cdot 4 = 15/8 = 2 - 1/8 = H(X)$

$\Rightarrow$ quindi φ_1 è ottima per la v.a. X (tra i codici u.d. con blocchi di lunghezza $n = 1$)

$\mathbb{E}[L^{(2)}] = 1/2 \cdot 1 + 4 \cdot 1/8 \cdot 3 = 1/2 + 1/2\, 3 = 2$ bit $= H(Y) \Rightarrow \varphi_2$ è ottima per Y.

Se uso φ_2 per X, ottengo $\mathbb{E}[L] = 1/2 \cdot 1 + \underbrace{1/4 \cdot 3 + 1/8 \cdot 3 + 2/16 \cdot 3}_{(1/2 \cdot 3)} = 1/2 + 3/2 = 2 > \mathbb{E}[L^{(1)}]$

$2 - 1/8 = H(X)$

$\Rightarrow \varphi_2$ non è ottima per X

Di quanto è maggiore $\mathbb{E}[L]$ di $H(X)$? (ridondanza)

$\mathbb{E}[L] = 2 \qquad H(X) = 2 - 1/8 \qquad \mathbb{E}[L] - H(X) = 1/8$

$\forall$ codifica per la v.a. X : $\mathbb{E}[L] \geq H(X)$ → d.d.p $p = (p_1, \dots, p_k)$ $(|\mathcal{B}| = 2)$

(Caso binario: $D = 2$)

$$q_i = \frac{2^{-\ell_i}}{\alpha} \qquad \alpha = \sum_{i=1}^{k} 2^{-\ell_i}$$

$$D(p\|q) = \sum_{i=1}^{k} p_i \log \frac{p_i}{2^{-\ell_i}} \cdot \alpha = \underbrace{\mathbb{E}[L] - \overbrace{H(X)}^{H(p)}}_{\text{ridondanza}} + \log \alpha$$

Se $\alpha = 1$, la ridondanza è $= D(p\|q)$

Se $\alpha \leq 1$, la ridondanza è $\geq D(p\|q)$

(2) Shannon-Fano : non è (in generale) ottima per n finito
→ Fano : non è (" ") ottima per n finito
Huffman : è ottima per n finito

a	13/100
b	10/100
c	3/100
d	15/100
e	18/100
f	41/100

$A = \{a, b, c, d, e, f\}$

$p = (0.13, 0.1, 0.03, 0.15, 0.18, 0.41)$

Fano :

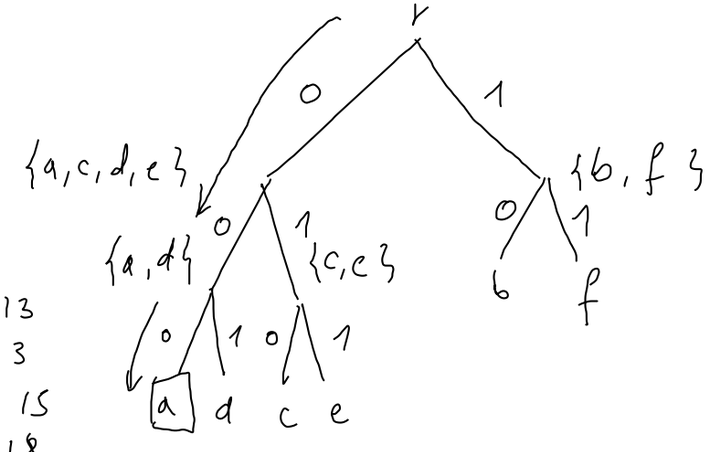
	$q(x)$	L
a	000	3
b	10	2
c	010	3
d	001	3
e	011	3
f	11	2

$\{f\}, \{a,b,c,d,e\} \rightarrow |41 - 59| = 18$

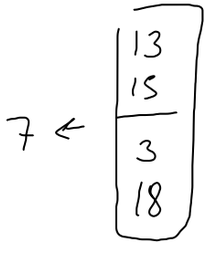
$\{c,f\}, \{a,b,d,e\} \rightarrow |44 - 56| = 12$

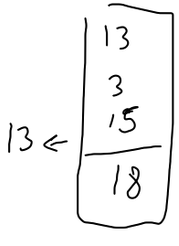
$\{a,c,d,e\}, \{b,f\} \rightarrow |49 - 51| = 2$

$E[L^{Fano}] = 2 \cdot (0.51) + 3 \cdot (0.49) = 2.49$



a	13
c	3
d	15
e	18

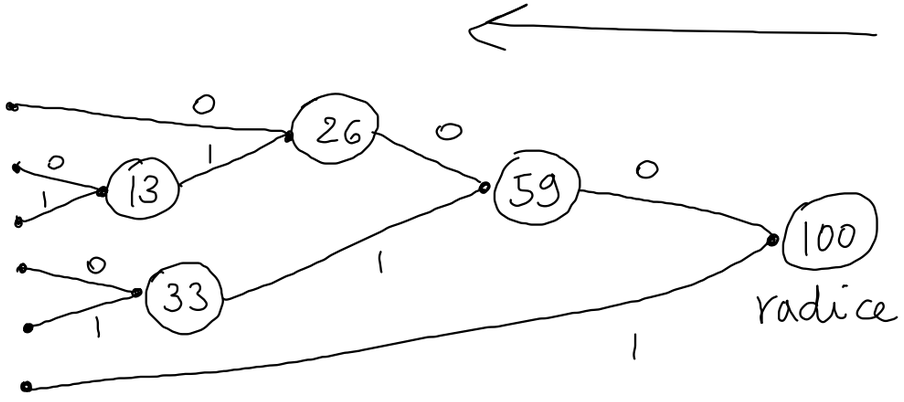




3

Huffman

a	13/100	(13)
b	10/100	(10)
c	3/100	(3)
d	15/100	(15)
e	18/100	(18)
f	41/100	41



φ	L
000	3
0010	4
0011	4
010	3
011	3
1	1

$$\mathbb{E}[L]_{\text{Huffman}} = 0.13 \cdot 3 + 0.10 \cdot 4 + \dots = 3 \cdot (0.13 + 0.15 + 0.18) + 4 \cdot (0.10 + 0.03) + 1 \cdot (0.41)$$

$$= 2.31 < \mathbb{E}[L^{\text{Fano}}] = 2.49.$$

$\Rightarrow$ la codifica di Fano non è ottima per questa d.d.p.

$|A| = K$
$\downarrow$
$|B| = D$

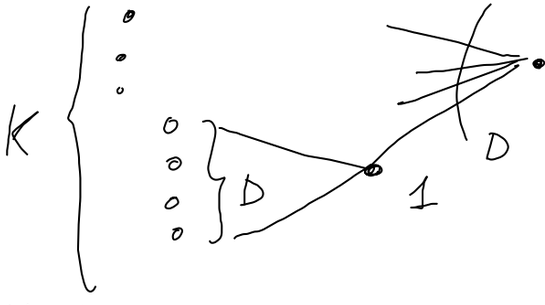
③ Codifica di Huffman per alfabeto di codifica D-ario con $D > 2$

Aggreghiamo ad ogni passo i D simboli meno probabili.

Dà un codice ottimo solo se l'albero di codice risultante è completo: ogni nodo ha o 0 figli o esattamente D figli; questo accade se e solo se $K \overset{(*)}{=} D + j \cdot (D-1)$ per qualche intero nonnegativo j.

$\iff K - j(D-1) = D$ per qualche j

$\iff K - D \equiv 0 \mod (D-1)$



Se la condizione (*) non è soddisfatta, aggiungiamo un numero minimo di simboli fittizi (a probabilità zero) in modo tale da soddisfarla.

Es. $A = \{a, b, c, d, e, f\}$ $(K = 6)$. Voglio $D = 3$.

$p = (6/18, 4/18, 3/18, 2/18, 2/18, 1/18)$

$K = 6$ $D = 3$ $\exists j$ intero t.c. $6 = 3 + j(2)$ $\rightarrow$ (*) non è soddisfatta

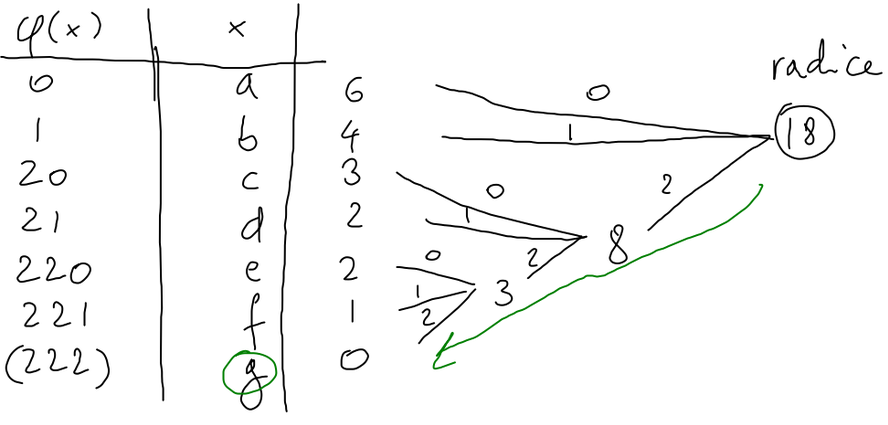
a	6		radice
b	4		
c	3	12	
d	2		
e	2	5	
f	1		

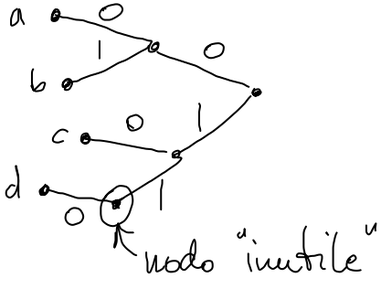
NON è ottimo per p

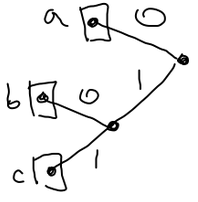
Aggiungo un simbolo fittizio g con prob. $0 = 0/18$.

$\Rightarrow$ $K = 7$ $\quad 7 \overset{?}{=} 3 + j(2)$ $\quad$ OK : $j = 2$

$\varphi(x)$	x
0	a
1	b
20	c
21	d
220	e
221	f
(222)	g





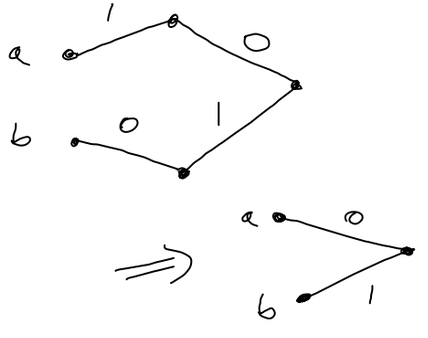


④ Quali di questi insiemi di parole di codice sono ottenibili tramite una codifica di Huffman?

(a) $\{0, 10, 11\}$: sì

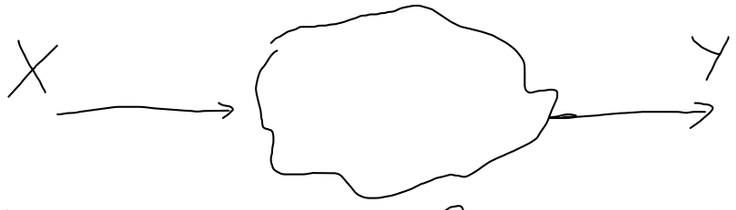
(b) $\{00, 01, 10, 110\}$: no : non è ottimo!
$\{00, 01, 10, 11\}$

(c) $\{01, 10\}$: no : non è ottimo



⑤ Decomprimere la seguente sequenza in base 10 compressa col metodo Ziv-Lempel LZ77 : $\boxed{605651600640}$ $\Longrightarrow$ decodifica ?

Calcolo della capacità di canale

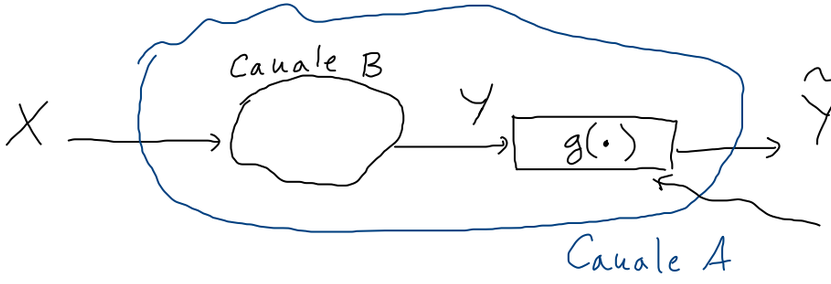


Capacità del canale: $\max_{p_X} I(X;Y) = C$

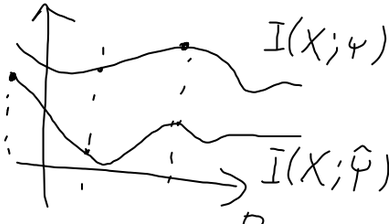
$p(y) = p(y|x) \cdot p(x) \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$

$\Gamma_{ij} = p(y_j|x_i)$ matrice

(1)



g è una funzione (deterministica)



Dimostrare che $\boxed{C_A \leq C_B}$.

$C_A = \max_{p_X} I(X;\tilde{Y}) \overset{?}{\leq} \max_{p_X} I(X;Y) = C_B$

$X \to Y \to \tilde{Y}$

Mostriamo che, qualunque sia p_X, ho $I(X;\tilde{Y}) \overset{?}{\leq} I(X;Y)$

2° teorema di elaborazione dati: se vale la catena di Markov $X \to Y \to Z$ allora ho $I(X;Y) \geq I(X;Z)$

Esercizi. Calcolo della capacità di canale.

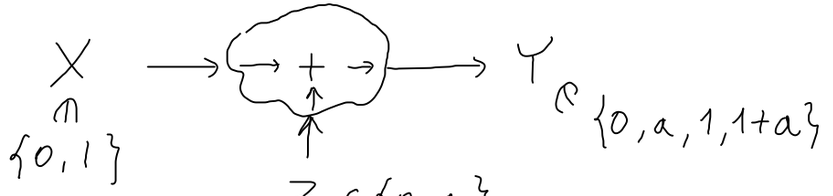
(1) Vale $X \to Y \to \tilde{Y}$ se la v.a. $\tilde{Y}$ è indipendente da X data Y

ovvero: $\tilde{Y}|Y$ è indipendente da X

Nel nostro caso vale perché $\tilde{Y} = g(Y)$

---

(2)



$a \in \mathbb{R}$ è un parametro

$Pr[Z=0] = 1/2$

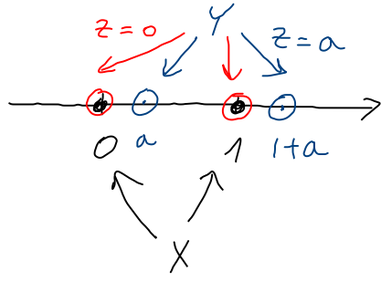
$Pr[Z=a] = 1/2$

X \ Y	Z→ 0	a
0	0	a
1	1	1+a

$Pr[X=0] =: \alpha$

$Pr[X=1] =: 1-\alpha$

per qualche $\alpha \in [0,1]$.



Svolgimento.

Caso $\boxed{a=0}$: $Y = X + 0 = X$;

capacità del canale: $\max_{p_X} I(X;Y) = \max_{p_X} \overbrace{I(X;X)}^{H(X)} = 1$ bit. ($p_X = (1/2, 1/2)$)

Caso $\boxed{a \neq 1 \ (\text{e } a \neq -1)}$

Capacità: $\max_{p_X} I(X;Y) = \max_{p_X} [H(Y) - H(Y|X)]$

$= \max_{p_X} [H(X) - H(X|Y)]$

$a = 1$

X \ Y	Z→ 0	1
0	0	1
1	1	2

Scenario : $a \neq 1$, $a \neq -1$ $(a \neq 0)$

$X \downarrow$ \ $Y \rightarrow$	0	a
0	0	a
1	1	1+a

4 valori tutti distinti

In questo caso,

$H(X|Y) = 0$ perché X è completamente determinata data Y.

$\Rightarrow$ capacità $= \max_{P_X} [H(X) - 0] = 1$ bit.

(si prende $P_X = (1/2, 1/2)$)

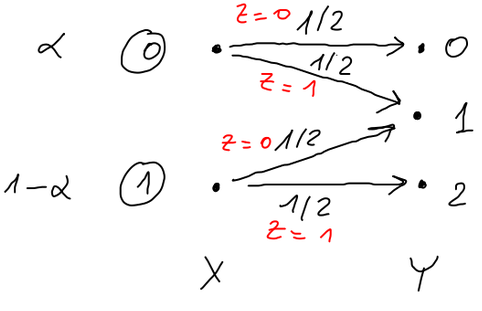
Ultimo caso: $\boxed{a = 1}$ oppure $\boxed{a = -1}$.

$X \downarrow$ \ $Y \rightarrow$	0	1	2
0	0	1	
1	1		2

$Pr[Y=0] = Pr[X=0 \wedge Z=0] = Pr[X=0]\,Pr[Z=0] = \alpha \cdot 1/2$

$Pr[Y=1] = Pr[X=0 \wedge Z=1] + Pr[X=1 \wedge Z=0] = \alpha \cdot 1/2 + (1-\alpha) \cdot 1/2 = 1/2$.

$Pr[Y=2] = Pr[X=1 \wedge Z=1] = (1-\alpha) \cdot 1/2$.



Canale con cancellazione con parametro $\varepsilon = 1/2$

Capacità del canale con cancellazione:

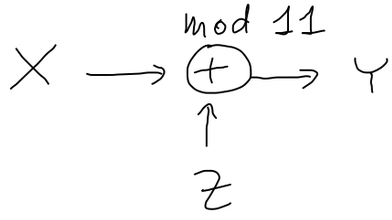
$C = 1 - \varepsilon = 1 - 1/2 = 1/2$ bit.

(3) $\mathcal{X} = \{0, 1, \ldots, 10\}$

$\mathcal{Z} = \{1, 2, 3\}$

$Y = (X + Z) \bmod 11$

$\mathcal{Y} = \{0, 1, \ldots, 10\}$



con:

$Pr[Z=1] = Pr[Z=2] = Pr[Z=3] = 1/3$.

(a) Trovare la capacità del canale

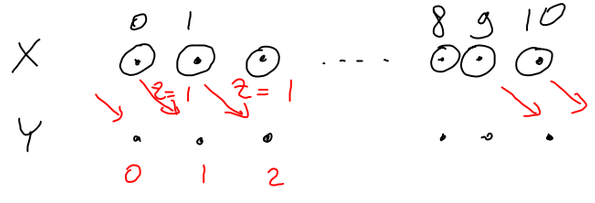
(b) Determinare la p_X corrispondente (quella che massimizza $I(X;Y)$)

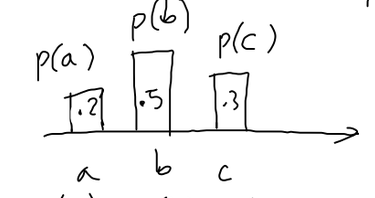
Sol.

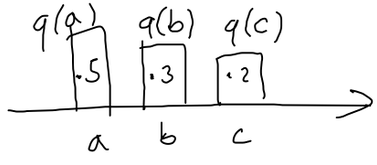
$$H(Y|X) = H((X+Z) \bmod 11 \mid X)$$
$$= H(Z|X)$$
$$= H(Z) = \log 3.$$

p_X

$$I(X;Y) = H(Y) - H(Y|X) =$$
$$= H(Y) - \log 3$$







$H(p) = H(q)$

$$C = \max_{p_X} \left[H(Y) - \log 3 \right] = \left(\max_{p_X} H(Y) \right) - \log 3$$

Certamente ho $H(Y) \leq \log |\mathcal{Y}| = \log 11$ ho =

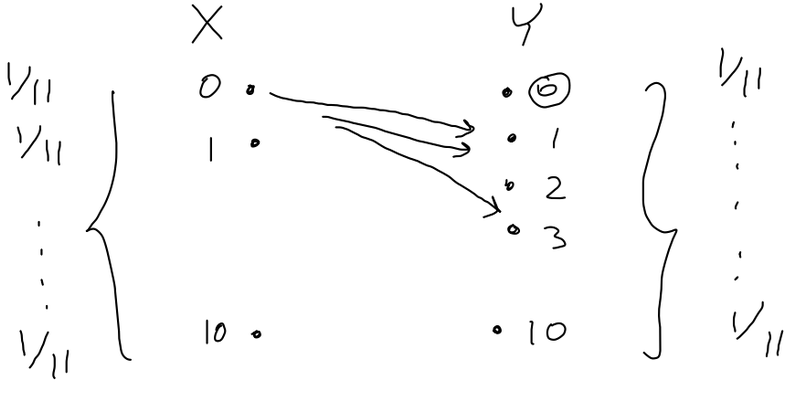
In effetti posso avere $H(Y) = \log 11$ quando p_Y è uniforme

Posso scegliere p_X tale da avere p_Y uniforme ? 0 1 10

Sì, basta scegliere p_X essa stessa uniforme : $p_X = (1/11, 1/11, \ldots, 1/11)$

$\rightarrow C = \log 11 - \log 3 = \log 11/3$. □

scelgo p_X uniforme (e quindi p_Y uniforme)

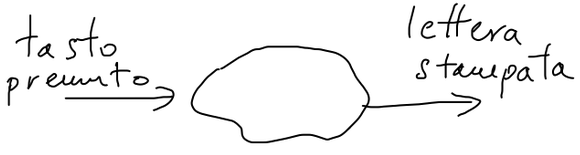


$3 \cdot (1/11 \cdot 1/3) = 1/11$

④ "Macchina da scrivere rumorosa" $\quad \mathcal{Y} = \mathcal{X} = \{x_1, x_2, \ldots, x_{26}\}$

Consideriamo una macchina da scrivere con 26 tasti 'A', 'B', 'C', ..., 'Z'

→ (a) Se la pressione di ogni tasto produce la corrispondente lettera, quanto vale la capacità?



$\mathcal{Y} = \mathcal{X}$

(b) se invece: 'A' $\xrightarrow{1/2}$ 'A', $\xrightarrow{1/2}$ 'B'; 'B' $\xrightarrow{1/2}$ 'B', $\xrightarrow{1/2}$ 'C'; ... 'Z' $\xrightarrow{1/2}$ 'Z', $\xrightarrow{1/2}$ 'A', quanto vale la capacità?

Sol. (a)

$$C = \max_{p_X} I(X;Y) = \max_{p_X} \left(H(X) - H(X|Y)\right)$$ ↙ p_X uniforme

In questo caso, $H(X|Y) = 0$; quindi $C = \max_{p_X} H(X) = \log|\mathcal{X}| = \log 26$.

(b) $C = \max_{p_X} I(X;Y) = \max_{p_X} \left(H(Y) - H(Y|X)\right)$

$$H(Y|X) = \sum_{i=1}^{26} p(x_i) \underbrace{H(Y|X=x_i)}_{\text{è l'entropia di } p_{Y|X=x_i}} = \sum_{i=1}^{26} p(x_i) = 1$$

$H(Y|X=x_i) = 1$

$X = x_7$ $\qquad$ $1/2 \quad 1/2$

$Y = ?$ $\qquad$ $Y \in \{x_7, x_8\}$

$p_{Y|X=x_7} = (1/2, 1/2)$

$= (0, 0, 0, \ldots, 1/2, 1/2, 0, \ldots, 0)$

$$C = \max_{P_X} \left(H(Y) - 1 \right) = \left(\max_{P_X} H(Y) \right) - 1$$

Ho $H(Y) \leq \log |\mathcal{Y}| = \log 26$

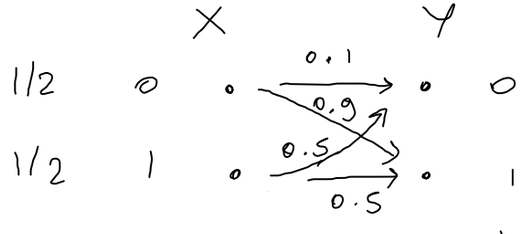
e $H(Y) = \log 26$ quando p_Y è uniforme ;

P_Y è uniforme se P_X è uniforme $\longrightarrow$ $\max_{P_X} H(Y) = \log 26$ (P_X è uniforme)

$$\longrightarrow C = (\log 26) - 1 = \log 26 - \log 2 = \log 26/2 = \log 13 . \square$$

=

$X \rightarrow \oplus \leftarrow Z$ $\quad Y = X + Z$



$Pr[Y=0] =$ $\quad Pr[Y=0|X=1]$

$= Pr[X=0] \cdot (0.1) + Pr[X=1] \cdot (0.5)$

$Pr[Y=0|X=0]$

$P_X \in \mathbb{R}^{|\mathcal{X}|}$

$P_Y \in \mathbb{R}^{|\mathcal{Y}|}$

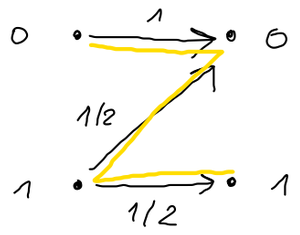
$$\Gamma = \left[\; p(y_j|x_i) \; \right] \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{Y}|}$$

$$p(y_j) = \sum_{i=1}^{K} p(y_j|x_i) \cdot p(x_i)$$

Se conosco P_X e conosco il canale (quindi conosco Γ) come determino P_Y ?

$$\boxed{P_Y^T = P_X^T \, \Gamma} .$$

⊙ Canale Z : Matrice di transizione $\Gamma = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$ ("0", "1")



(Se $p_X = (1/2, 1/2)$, $p_Y = p_X \Gamma = \frac{1}{2}(1,0) + \frac{1}{2}(1/2, 1/2)$
$= (3/4, 1/4)$)

$$C = \max_{p_X} I(X;Y) = \max_{p_X} [H(Y) - H(Y|X)]$$

$p_X = (1-\alpha, \alpha)$ con $\alpha \in [0,1]$ (0 1)

$$\rightarrow H(Y|X) = Pr[X=0] H(Y|X=0) + Pr[X=1] H(Y|X=1)$$
$$= (1-\alpha) H(Y|X=0) + \alpha H(Y|X=1)$$
$$= (1-\alpha) \cdot 0 + \alpha \cdot \underbrace{H((1/2, 1/2))}_{=1} = \alpha$$

$$p_Y = p_X \Gamma = (1-\alpha)(1,0) + \alpha (1/2, 1/2)$$
$$= (1 - \alpha + \alpha/2, \alpha/2)$$
$$= (1 - \alpha/2, \alpha/2)$$
$H((1-\varepsilon, \varepsilon)) = h_2(\varepsilon)$

$$\rightarrow H(Y) = h_2(\alpha/2)$$
$$C = \max_{p_X} [h_2(\alpha/2) - \alpha] = \max_{0 \le \alpha \le 1} [h_2(\alpha/2) - \alpha]$$

$$C = \max_{0 \le \alpha \le 1} \left[h_2(\alpha/2) - \alpha \right] = \max_{0 \le \alpha \le 1} f(\alpha)$$

$$f(\alpha) = h_2(\alpha/2) - \alpha$$

$$f(0) = 0 - 0 = 0$$

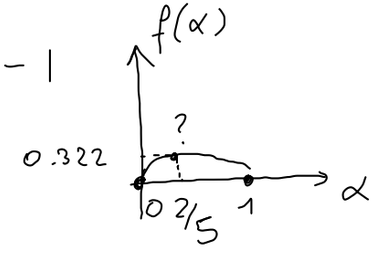
$$f(1) = 1 - 1 = 0$$

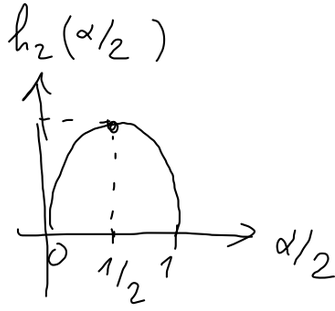
$$f'(\alpha) = \left[-\alpha/2 \log \alpha/2 - (1-\alpha/2) \log (1-\alpha/2) - \alpha \right]'$$

$$= -1/2 \log \alpha/2 - \cancel{\alpha/2 \cdot 2/\alpha \cdot 1/2} + 1/2 \log (1-\alpha/2) + \cancel{(1-\alpha/2)} \frac{1}{\cancel{1-\alpha/2}} \cdot 1/2 - 1$$

$$= -1/2 \log \frac{\alpha}{2} + 1/2 \log (1-\alpha/2) - 1$$

$$= 1/2 \log \frac{1-\alpha/2}{\alpha/2} - 1$$





$$f'(\alpha) = 0 \iff \log_2 \frac{1-\alpha/2}{\alpha/2} = 2$$

$$\iff \frac{1-\alpha/2}{\alpha/2} = 4$$

$$1 - \alpha/2 = 4 \cdot \alpha/2$$

$$5\,\alpha/2 = 1$$

$$\alpha = 2/5$$

$$f(2/5) = h_2(1/5) - 2/5$$

$$\approx 0.322 \text{ bit} \quad . \quad \square$$

① Distanza di Hamming

n-ple (sequenze di lungh. n su uno stesso alfabeto A)

$x = (x_1, x_2, \dots, x_n) \in A^n$
$\quad\quad \}=? \quad \}=? \quad\quad \}=?$
$y = (y_1, y_2, \dots, y_n) \in A^n$

Distanza di Hamming: $d_H(x,y) = \#\{\text{posizioni } i \ (1 \le i \le n) \text{ tale che } x_i \neq y_i\}$

Sfera di Hamming: $S_\rho(x)$ sfera di Hamming di raggio ρ e centro x

Def: $S_\rho(x) = \{ y \in A^n : d_H(x,y) \le \rho \}$.

Esempio: $A = \{0,1\}$, $n = 3$

$A^n = \{000, 001, 010, 100, 101, 110, 011, 111\}$

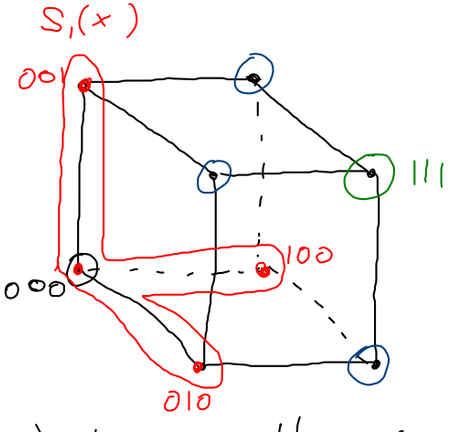
$x = 000$

$S_0(x) = \{x\} = \{000\}$

$S_1(x) = \{000\} \cup \{001, 010, 100\}$

$S_2(x) = S_1(x) \cup \{011, 101, 110\}$

$S_3(x) = A^3$

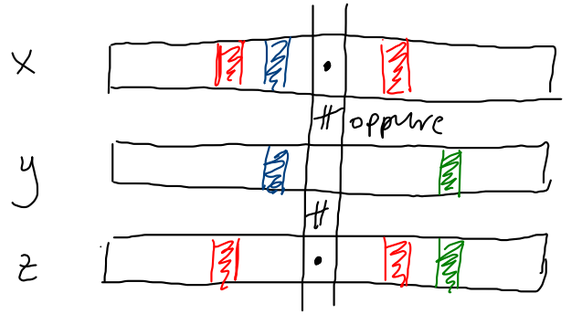


(Iper)cubo di Hamming A^n

$d_H(\cdot,\cdot)$ è una distanza:

1. $d_H(x,y) \geq 0$ e $d_H(x,y) = 0 \iff x = y$ $(\forall x\, \forall y)$ (definita positiva) ✓
2. $d_H(x,y) = d_H(y,x)$ $(\forall x\, \forall y)$ (simmetrica) ✓
3. $d_H(x,z) \leq d_H(x,y) + d_H(y,z)$ $(\forall x\, \forall y\, \forall z \in A^n)$?

(disuguaglianza triangolare)



Proposizione. Per ogni n-pla $x \in \{0,1\}^n$ e ogni $0 \leq p \leq 1/2$, ho $|S_{\rho n}(x)| \leq 2^{h_2(p)\cdot n}$. $\rho = p \cdot n$

$S_{\rho n}(x) \subseteq A^n$

$|S_{\rho n}(x)| \leq |A^n| = 2^n$

$$|S_{pn}(x)| \leq 2^{h_2(p)\cdot n} \qquad (A=\{0,1\})$$

$$\forall n \in \mathbb{N} \quad \forall x \in A^n \quad \forall p \in [0, 1/2].$$

Dim.

$$1 = (p+(1-p))^n =$$

$$= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \geq$$

$$\geq \sum_{i=0}^{pn} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=0}^{pn} \binom{n}{i} (1-p)^n \left(\frac{p}{(1-p)}\right)^i \geq$$

$$\geq \sum_{i=0}^{pn} \binom{n}{i} (1-p)^n \left(\frac{p}{(1-p)}\right)^{pn}$$

$$= (1-p)^n \sum_{i=0}^{pn} \binom{n}{i} p^{pn} (1-p)^{-pn}$$

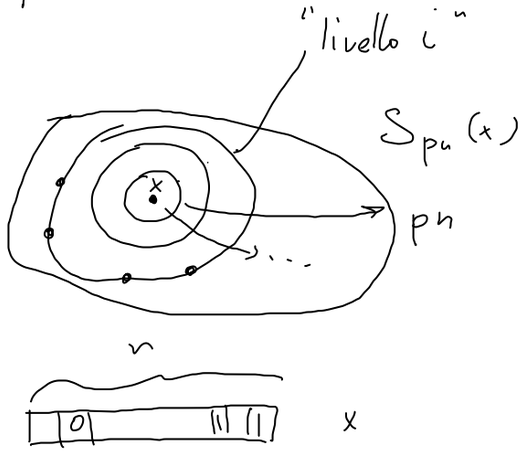
$$= \sum_{i=0}^{pn} \binom{n}{i} p^{pn} (1-p)^{(1-p)n}$$

$\beta^j \geq \beta^k$ se $\beta \leq 1$ e $k \geq j$

$p \in [0, 1/2] \Rightarrow$

$p \leq 1/2 \Rightarrow p/(1-p) \leq 1$

$1-p \geq 1/2$



$$|S_{pn}(x)| = \sum_{i=0}^{pn} |\text{livello } i|$$

Se $y \in$ livello i di $S_{pn}(x)$

i posizioni in cui y differisce da x

Ci sono $\binom{n}{i}$ stringhe y nel livello i.

$$1 \geq \sum_{i=0}^{pn} \binom{n}{i} p^{pn} (1-p)^{(1-p)n}$$

indipendenti dall'indice i

$$\binom{n}{i} = |\text{livello } i|$$

$$= p^{pn}(1-p)^{(1-p)n} \sum_{i=0}^{pn} \binom{n}{i}$$

= somma delle cardinalità dei primi pn livelli della sfera di Hamming $S_{pn}(x)$

$$= |S_{pn}(x)|$$

$$= 2^{pn \log p} \cdot 2^{(1-p)n \log(1-p)} \cdot |S_{pn}(x)|$$

$$= 2^{n[p \log p + (1-p)\log(1-p)]} \cdot |S_{pn}(x)|$$

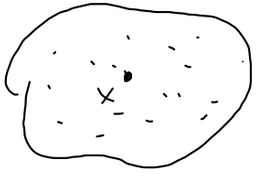
$$= 2^{-n h_2(p)} |S_{pn}(x)|$$

$$p = 2^{\log p}$$

$$(1-p) = 2^{\log(1-p)}$$

$\Rightarrow$ Moltiplicando per $2^{n h_2(p)}$, ottengo

$$|S_{pn}(x)| \leq 2^{n h_2(p)}.$$



$\forall p \in [0, 1/2]$, $x \in \{0,1\}^n$.

Per sequenze non necessariamente binarie, (A non necess. $\{0,1\}$)

vale comunque: $|S_{pn}(x)| \leq (n+1)\, 2^{n h_2(p)}$

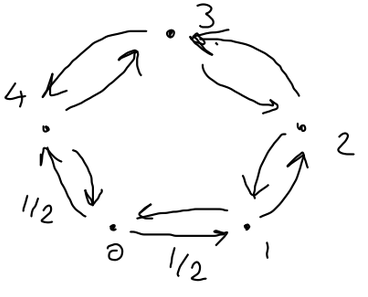
Equivalentemente, per $\rho = p^n$, $\quad |S_\rho(x)| \leq (n+1)\, 2^{n h_2(n/\rho)}$

$p = n/\rho$

---

② Canale con alfabeto di ingresso e di uscita $A = \{0,1,2,3,4\} = X = Y$

$$p(y|x) = \begin{cases} 1/2 & \text{se } y = x \pm 1 \bmod 5 \\ 0 & \text{altrimenti} \end{cases}$$



Calcolare la capacità del canale.

Matrice di transizione: $\Gamma =$

	0	1	2	3	4
0	0	1/2	0	0	1/2
1	1/2	0	1/2	0	0
2	0	1/2	0	1/2	0
3	0	0	1/2	0	1/2
4	1/2	0	0	1/2	0

È un canale simmetrico.

$\Rightarrow$ la capacità $C = \max_{p_X} \left[H(Y) - \underbrace{H(Y|X=0)}_{\text{non dipende da } p_X} \right] = \left(\max_{p_X} H(Y) \right) - H(Y|X=0)$

$$H(Y|X=0) = H\left((0, 1/2, 0, 0, 1/2)\right) = h_2(1/2) = 1 .$$

$$\max_{p_X} H(Y) \leq \log |\mathcal{Y}| = \log 5$$

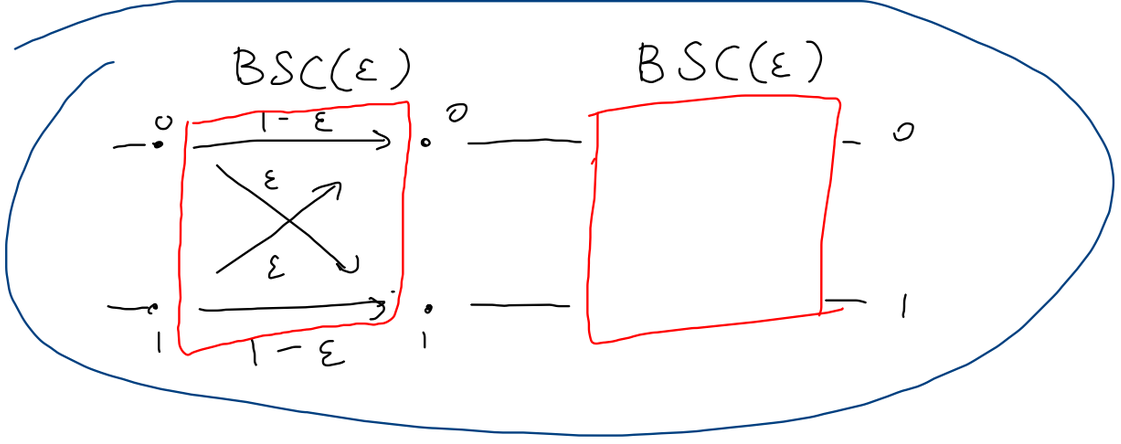
↑ Ho uguaglianza se e solo la p_Y è uniforme su $\mathcal{Y}$

Ho p_Y uniforme se la p_X è uniforme:

$$\underbrace{(1/5, 1/5, 1/5, 1/5, 1/5)}_{p_X} \lceil = \underbrace{(1/5, 1/5, 1/5, 1/5, 1/5)}_{p_Y}$$

$\Rightarrow \max_{p_X} H(Y) = H\left((1/5, 1/5, \ldots, 1/5)\right) = \log 5 .$ $\Rightarrow C = (\log 5) - 1 .$

□



Qual è la capacità della cascata (concatenazione) di 2
(o di $n \geq 2$) canali binari simmetrici con parametro ε.

Classificazione Bayesiana

Esempio. Specie di piante : $\{$Asteraceae, Orchidaceae, Rubiaceae$\} = \mathcal{Y}$ (classi "cause")

Colori dei fiori : $\{$viola, bianco, giallo$\} = \mathcal{X}$ (valori osservabili)

100 osservazioni:

- 40 sono Asteraceae, di cui 5 con fiori viola, 20 con fiori bianchi, 15 con fiori gialli
- 40 sono Orchidaceae, di cui 20 " " viola, 10 " " bianchi, 10 " " gialli
- 20 sono Rubiaceae, di cui tutte con fiori gialli

Un classificatore è una funzione $h : \mathcal{X} \to \mathcal{Y}$

Data un'osservazione $x \in \mathcal{X}$, vorremmo che $h(x)$ massimizzasse la probabilità di corretta classificazione : $\Pr[h(X) = Y \mid X = x]$

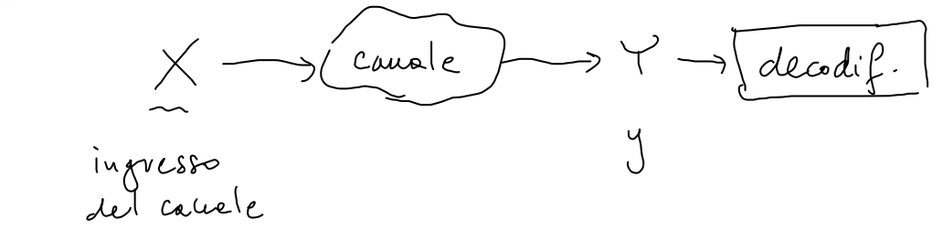
Il classificatore bayesiano associa ad ogni $x \in \mathcal{X}$ la classe $\hat{y} \in \mathcal{Y}$ che massimizza tale probabilità :

(X, Y)

$$\hat{y} = \underset{y_j \in \mathcal{Y}}{\operatorname{argmax}} \Pr[\hat{y} = Y \mid x]$$

Esercizi. Classificazione bayesiana. Esempio di decodifica Ziv-Lempel. Concatenazione di canali.

Codici di canale:



Criterio della decodifica a massima verosim.:

data un'uscita del canale $y \in \mathcal{Y}$,

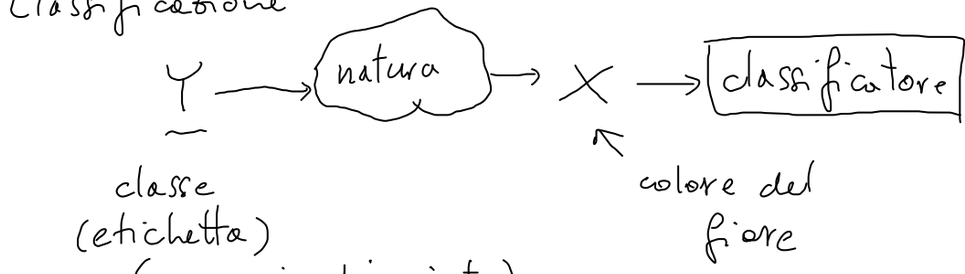
restituisci la parole di codice $\hat{x}$

che massimizza $p(\hat{x}|y)$:

$$\hat{x} = \underset{x_i}{\operatorname{argmax}} \; p(x_i|y)$$

fisso (dato)

Classificazione



Classific. bayesiana,

data l'osservazione $x \in \mathcal{X}$,

restituisci la classe (etichetta) $\hat{y} \in \mathcal{Y}$

che massimizza $p(\hat{y}|x)$:

$$\hat{y} = \underset{y_j}{\operatorname{argmax}} \; p(y_j|x) .$$

$p(x,y)$	viola	bianchi	gialli
Asteraceae	5%	20%	15%
Orchidaceae	20%	10%	10%
Rubiaceae	0%	0%	20%

(y: Asteraceae, Orchidaceae, Rubiaceae; x: viola, bianchi, gialli)

$p(y\|x)$	viola	bianchi	gialli
Asteraceae	1/5	2/3	3/9
Orchidaceae	4/5	1/3	2/9
Rubiaceae	0	0	4/9
	1/5	1/3	5/9

(1/4 del totale)

$$p(y|x) = \frac{p(x,y)}{p(x)} \qquad \frac{5}{25}, \frac{20}{25}, \frac{0}{25}$$

Quindi $h(\text{viola}) = \text{Orchidaceae}$ è l'ipotesi bayesiana per l'osservazione $x = \text{viola}$

$h(\text{bianco}) = \text{Asteraceae}$ " " " " $x = \text{bianco}$

$h(\text{giallo}) = \text{Rubiaceae}$ " " " " $x = \text{giallo}$

Qual è la prob. di errore complessiva?

$p(X = \text{viola}) = 25/100 = 1/4$

$p(X = \text{bianco}) = 30/100 = 3/10$

$p(X = \text{giallo}) = 45/100 = 9/20$

Prob. di errata classificazione $= 1/4 \cdot 1/5 + 3/10 \cdot 1/3 + 9/20 \cdot 5/9 = 8/20 = 40\%$.

① Consideriamo due v.a. X_1 e X_2.

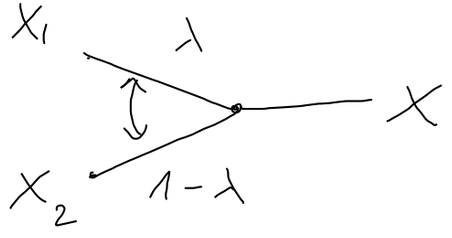
Definiamo una terza v.a. $X = \begin{cases} X_1 & \text{con prob. } \lambda \\ X_2 & \text{con prob. } 1-\lambda \end{cases}$

(λ parametro fissato, $\lambda \in [0,1]$)

Dimostrare che $H(X) \geq \lambda H(X_1) + (1-\lambda) H(X_2)$.

Definisco una v.a. Z binaria:

$$Z = \begin{cases} 1 & \text{se } X = X_1 \quad (\text{prob. } \lambda) \\ 2 & \text{se } X = X_2 \quad (\text{prob. } 1-\lambda) \end{cases}$$



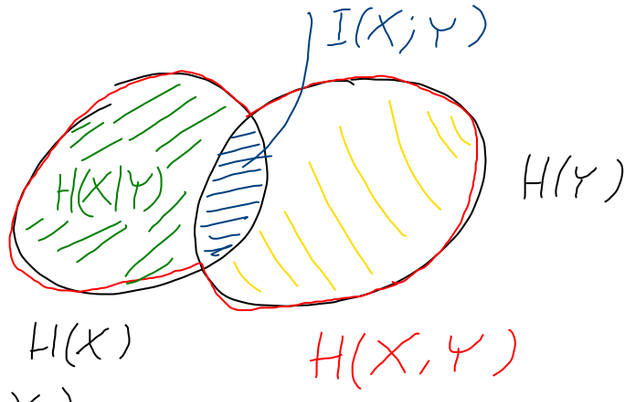
$$H(X) \geq H(X|Z) = \underbrace{Pr[Z=1]}_{\lambda} \underbrace{H(X|Z=1)}_{H(X_1)} + \underbrace{Pr[Z=2]}_{1-\lambda} \underbrace{H(X|Z=2)}_{H(X_2)} .$$

↑ aggiungere un condizionamento non può aumentare l'entropia di X

↑ proprietà entropia condizionata

(2) Consideriamo due v.a. X e Y.

Sia $\rho = \dfrac{I(X;Y)}{H(X,Y)}$



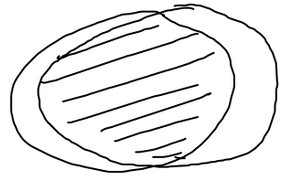
Dimostrare che:

$$H(X,Y) = H(X|Y) + I(X;Y) + H(Y|X)$$

(a) $0 \le \rho \le 1$

(b) Dare una condiz. necess. e suff. affinché $\rho = 0$

(c) " " " " " " " $\rho = 1$



$H(X) = H(Y) = H(X,Y) = I(X;Y)$

(a) $H(X,Y) > 0$ e $I(X;Y) \ge 0 \Rightarrow \rho \ge 0$

(* a meno che X = cost., Y = cost.)

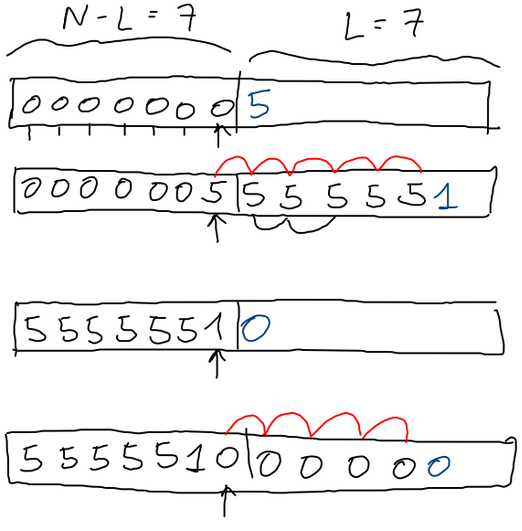
$I(X;Y) \le H(X,Y) \Rightarrow \rho \le 1$

(b) $\rho = 0 \iff I(X;Y) = 0 \iff X$ e Y sono v.a. indipendenti

→ (c) $\rho = 1 \iff I(X;Y) = H(X,Y) \iff H(X|Y) = 0$ e $H(Y|X) = 0 \iff$ La X è una funz. di Y
La Y è una funz. di X
X e Y sono in corrispondenza biunivoca.

③ Decodifica di Ziv-Lempel (LZ77) della stringa

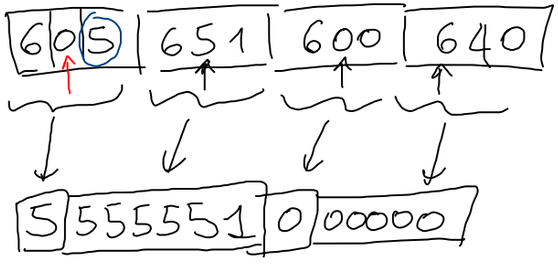
605 651 600 640 in alfabeto decimale ($N = 14$, $L = 7$, $K = 10$)



Lunghezza dei blocchi:

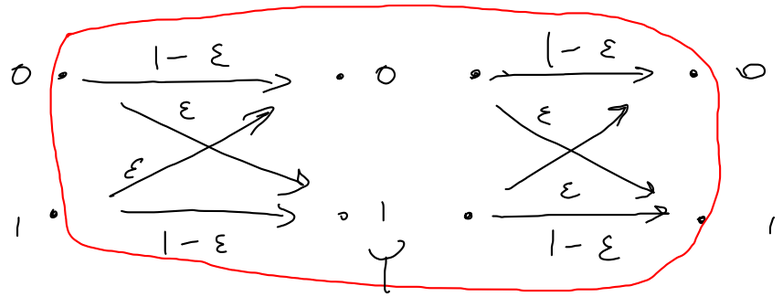
- Parte puntatore: $\lceil \log_{10}(N-L) \rceil = \lceil \log_{10} 7 \rceil = \underline{1}$
- Parte lunghezza: $\lceil \log_{10} L \rceil = \lceil \log_{10} 7 \rceil = \underline{1}$
- 1 simbolo aggiuntivo: $\underline{1}$

Totale: 3



Sequenza decodificata è

④ Concatenazione di canali $BSC(\varepsilon)$



Scrivere la matrice di transizione del canale ottenuto concatenando due canali $BSC(\varepsilon)$.

Matrice del primo canale : $\Gamma = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$

Matrice del secondo canale : $\Gamma = \;''$

Matrice di transizione del canale concatenato

$p_Y = p_X \Gamma$ (p_Y: vettore, p_X: riga)

$p_Z = p_Y \cdot \Gamma \quad \Rightarrow \quad p_Z = (p_X \Gamma) \cdot \Gamma = p_X \Gamma^2$

$$\Gamma^2 = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}\begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix} = \begin{pmatrix} (1-\varepsilon)^2+\varepsilon^2 & 2\varepsilon(1-\varepsilon) \\ 2\varepsilon(1-\varepsilon) & (1-\varepsilon)^2+\varepsilon^2 \end{pmatrix} \cdot \square$$

$$= \begin{pmatrix} \frac{1}{2}(1+(1-2\varepsilon)^2) & \frac{1}{2}(1-(1-2\varepsilon)^2) \\ \frac{1}{2}(1-(1-2\varepsilon)^2) & \frac{1}{2}(1+(1-2\varepsilon)^2) \end{pmatrix}$$

In generale, per induzione su n,

$$\Gamma^n = \begin{pmatrix} \frac{1}{2}(1+(1-2\varepsilon)^n) & \frac{1}{2}(1-(1-2\varepsilon)^n) \\ \frac{1}{2}(1-(1-2\varepsilon)^n) & \frac{1}{2}(1+(1-2\varepsilon)^n) \end{pmatrix}$$

$\xrightarrow[(\text{se } \varepsilon \in (0,1/2))]{n \to \infty}$ $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ Canale inutile.

① Il codice binario (su $GF(2)$) $C = \{000, 010, 100, 111\}$ ($n = 3$)

Il codice è lineare?

parole di codice

$n = 3$

$V^{(n)} = A^n = \{0,1\}^3$ ($A = \{0,1\}$)

$a_1 a_2 a_3 \in V^{(3)}$

$b_1 b_2 b_3 \in V^{(3)}$

$\rightarrow (a_1 + b_1, a_2 + b_2, a_3 + b_3) \in V^{(3)}$

$|C| = |V^{(k)}|$ per qualche k intero ≥ 0

$= 2^k$ → in questo ok per $\boxed{k = 2}$

Devo verificare che $\forall x \in C$,

$0 \cdot x \in C$ → $0 \cdot x = 00\ldots0 \in C$

$1 \cdot x \in C$ → $1 \cdot x = x \in C$

$\forall x, x' \in C$, $x + x' \in C$

$x = 010$, $x' = 100$ → $x + x' = 110 \notin C$ → il codice non è lineare.

② Il codice binario $\mathcal{C} = \{0000, 0101, 1010, 1111\}$ è lineare? ($n = 4$)

$V^{(n)} = \{0,1\}^4$ $\quad |\mathcal{C}| = 4 = 2^2$ → Se $\mathcal{C}$ è lineare, la sua dimensione (dim. di $V^{(k)}$) è 2 (k)

$\mathcal{C} \ni 0000 + x = x \in \mathcal{C}$

$0101 + 0101 = 0000 \in \mathcal{C}$

$0101 + 1010 = 1111 \in \mathcal{C}$

$1111 + 0101 = 1010 \in \mathcal{C}$

$1111 + 1010 = 0101 \in \mathcal{C}$

$$G = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Matrice generatrice

$\mathcal{C} \ni x = u \cdot G$

- G è una matrice a elementi in $GF(2)$ di dimensione $k \times n \rightarrow 2 \times 4$
- Ciascuna riga di G deve essere una parola di codice
- Il rango di G deve essere k

$$G' = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$2^k \times n$

$$\begin{bmatrix} 0000 \\ 0101 \\ 1010 \\ 1111 \end{bmatrix}$$

$GF(q)$ esiste se e solo se $q = p^s$ con p primo, $s \geq 1$ intero

~~$GF(6)$~~

$GF(2)$, $GF(3)$, $GF(4)$

In $GF(p)$ con p primo: $GF(p) = \{0, 1, \dots, p-1\}$

- l'addizione è l'addizione modulo p : $(\mathbb{Z}_p, +)$ è un gruppo
- la moltiplicazione è la moltiplicazione modulo p

$GF(3)$

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

: $(\mathbb{Z}_p^*, *)$ è un gruppo

$\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$

GF(4)

$\{0, 1, \alpha, 1+\alpha\}$

$\alpha = 2$? NO

$1 + 1$

Non posso avere ciò

$0 \neq 0$

$x \cdot x' = 0$ Moltiplico per x^{-1}

$x' = x^{-1} \cdot 0 = 0$

$1 + 1 = 0$

+	0	1	α	$1+\alpha$
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

*	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

$x \cdot x' = y$

$\mathbb{Z}_4^*$

③ Il codice ternario ($\equiv$ in $GF(3)$) $\mathcal{C} = \{\overbrace{000000}^{n=6}, 012112, 021221\}$ è lineare?

$V^{(n)} = \{0,1,2\}^6$ $\quad |\mathcal{C}| = 3 = |V^{(k)}| = 3^k \Rightarrow$ dobbiamo avere $k = 1$

$C(n,k)$

$C(6,1)$

$\forall x \in \mathcal{C}$, $\forall a \in \{0,1,2\}$ $\quad a \cdot x \in \mathcal{C}$

$a \cdot 000000 = 000000 \in \mathcal{C}$
$a \cdot 012112$
$a \cdot 021221$

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

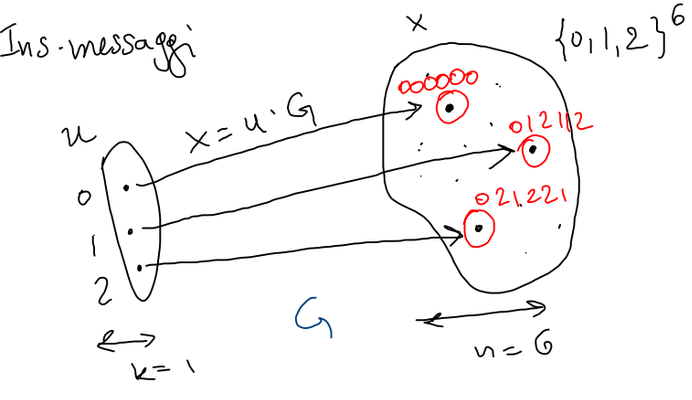
Se $a = 1$ $\quad 1 \cdot 012112 = 012112 \in \mathcal{C}$
$1 \cdot 021221 = 021221 \in \mathcal{C}$

$a = 2$ $\quad 2 \cdot 012112 = 021221 \in \mathcal{C}$
$2 \cdot 021221 = 012112 \in \mathcal{C}$

} ok

$\forall x, x' \in \mathcal{C}$, $x + x' \in \mathcal{C}$

$012112 + 021221 = 000000 \in \mathcal{C}$
$012112 + 012112 = 2 \cdot 012112 \in \mathcal{C}$
$021221 + 021221 = 2 \cdot 021221 \in \mathcal{C}$



→ $\mathcal{C}$ è lineare.

$$\begin{bmatrix} \cancel{000000} \\ 012112 \\ \cancel{021221} \end{bmatrix}$$

u
$\Leftrightarrow$
$k = 1$

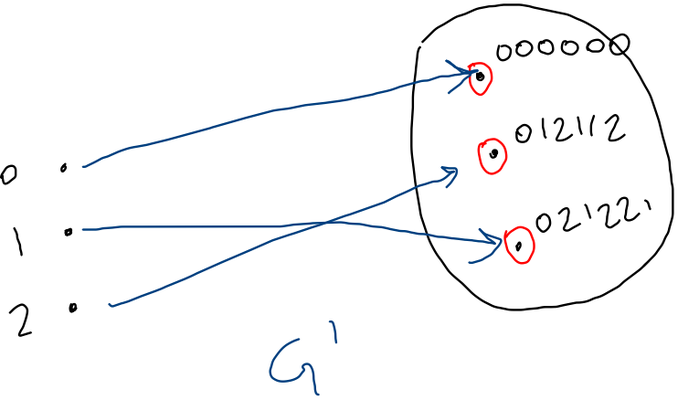
$(k \times n)$

$$G = [0\ 1\ 2\ 1\ 1\ 2] \in \mathbb{F}_3^{1\times 6}$$

G ha rango $1 = k$ ← x

u	
0	$0 \cdot 012112 = 000000$
1	$1 \cdot 012112 = 012112$
2	$2 \cdot 012112 = 021221$

Se invece avessi scelto $G' = [0\ 2\ 1\ 2\ 2\ 1]$



0	$0 \cdot 021221 = 000000$
1	$1 \cdot 021221 = 021221$
2	$2 \cdot 021221 = 012112$

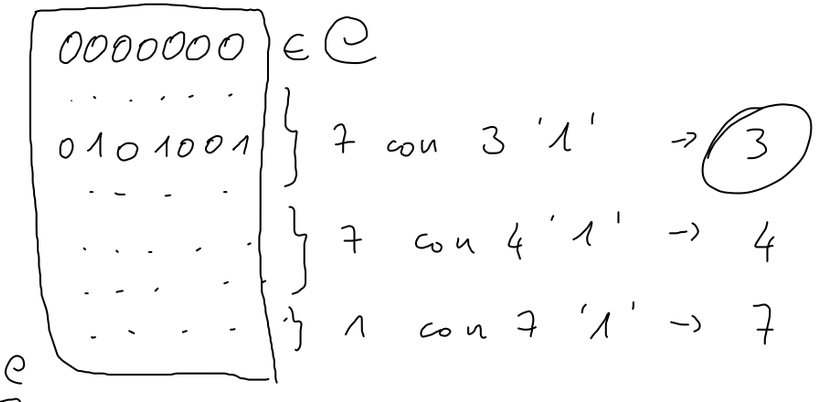
④ Un codice binario di tipo $C(7,4)$ (→ lineare) ha:

- 7 parole contenenti 3 simboli '1'
- 7 parole " 4 simboli '1'
- 1 parola " 7 simboli '1'

(a) Se il codice è usato solo per rilevare errori, quanti errori può rilevare?

(b) Qual è la probabilità di mancata rilevazione di errore nel canale BSC(ε)?

(a) $n=7$, $k=4$, $|\mathcal{C}| = 2^4 = 16$

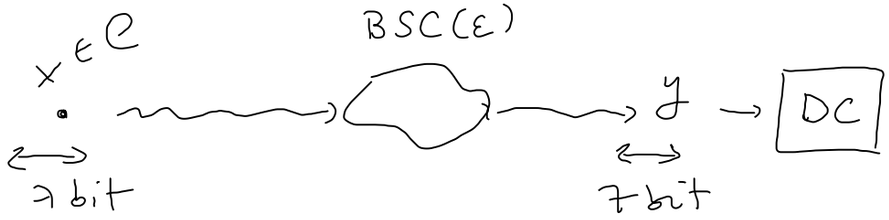


$$\#\text{errori rilevabili} = d_{min} - 1$$

$$= \min_{\substack{x \neq x' \\ x, x' \in \mathcal{C}}} d_H(x, x') - 1$$

$$= \min_{e \neq \vec{0}} \underbrace{d_H(x - x', 0000000)}_{e \in \mathcal{C}} - 1 = \min_{\substack{e \neq \vec{0} \\ e \in \mathcal{C}}} wt(e) - 1$$

$$= 3 - 1 = 2$$



Vari casi a seconda di $d_H(x,y)$:

- Se $d_H(x,y)=0$, $y=x \in \mathcal{C}$: nessun errore rilevato

$$(1-\varepsilon)^7$$

- Se $d_H(x,y)=1$; $y \notin \mathcal{C}$ rilevo errore

$$\binom{7}{1}\varepsilon^1(1-\varepsilon)^6 = 7\varepsilon(1-\varepsilon)^6$$

- Se $d_H(x,y)=2$; $y \notin \mathcal{C}$ rilevo errore

$$\binom{7}{2}\varepsilon^2(1-\varepsilon)^5$$

Scenari in cui l'errore (se c'è) è rilevato

Se $d_H(x,y) \geq 3$

y potrebbe appartenere a $\mathcal{C}$

(non riesco a rilevare errore)

$$\binom{7}{3}\varepsilon^3(1-\varepsilon)^4 + \binom{7}{4}\varepsilon^4(1-\varepsilon)^3 + \binom{7}{5}\varepsilon^5(1-\varepsilon)^2 + \binom{7}{6}\varepsilon^6(1-\varepsilon) + \binom{7}{7}\varepsilon^7 = O(\varepsilon^3)$$

(5) Un codice lineare binario ha matrice generatrice:

$n=2$

$$G = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \quad k=2$$

(1) Quante e quali sono le parole di codice? 4

(2) Qual è la distanza minima del codice? 6

(1) $V^{(n)} = \{0,1\}^9$ $n=9$ $k=2$ $V^{(k)} = \{0,1\}^2$ $|\mathcal{C}| = |V^{(k)}| = 2^2 = 4$

→ 4 parole di codice; quali sono?

$$\begin{matrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \\ x^{(4)} \end{matrix} \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad \begin{matrix} 6 \\ 6 \\ \cancel{0} \\ 6 \end{matrix}$$

$$d_{min} = \min_{e \neq \vec{0}} wt(e) = 6$$

$01\{101101 \in \mathcal{C}$

$101101\{01 \in \mathcal{C}$

⑥ Data una matrice $k \times n$ G in $GF(2)$

che genera un codice lineare $C = \{ u \cdot G \mid u \in \{0,1\}^k \} \subseteq \{0,1\}^n$

($u \cdot G$: n; u: k)

Mostrare che è sempre possibile

trovare una matrice H di dimensioni $n \times (n-k)$

tale che $C = \{ x \in \{0,1\}^n : x \cdot H = \vec{0} \}$.

Dim. $y \in C^\perp$: $\boxed{x_1 y_1 + x_2 y_2 + \dots + x_n y_n = 0}$ $\forall x \in C$

$\dim(C^\perp) + \underbrace{\dim(C)}_{k} = \dim(V^{(n)}) = n \rightarrow \dim(C^\perp) = n - k$

Prendo una qualunque base di $C^\perp$: ottengo $n-k$ vettori di lunghezza n linearmente indipendenti

Chiamiamoli $y^{(1)}, y^{(2)}, \dots, y^{(n-k)} \in GF(2)^n$

$\rightarrow$ ottengo $x \cdot H = \vec{0} = [0\ 0\ \dots\ 0]$

$H = \left[\left(y^{(1)}\right) \left(y^{(2)}\right) \left(y^{(3)}\right) \left(\ \right) \left(\ \right) \dots \left(y^{(n-k)}\right) \right]$

Es. (6.8 del libro)

Si consideri il codice (lineare) binario $C(5,1)$ con matrice generatrice $G = [1\ 1\ 1\ 1\ 1]$

($\mathbb{F}_2^n$, n, k, $(\mathbb{F}_2^n, +)$)

Costruire le parole del codice e una tabella di Slepian.

$k=1 \rightarrow$ I possibili messaggi sono in $\mathbb{F}_2^k = \mathbb{F}_2 = \{0,1\}$

$x = u \cdot G \rightarrow u=0 \rightarrow x=[0\ 0\ 0\ 0\ 0]$, $u=1 \rightarrow x=[1\ 1\ 1\ 1\ 1]$

Le parole di codice sono 00000 e 11111; è un codice a ripetizione di lunghezza 5.

		Parole di codice	Laterale	Sindromi
$\binom{5}{0}=1$	peso 0	00000	11111	0000^T
$\binom{5}{1}=5$	peso 1	00001	11110	0001^T
		00010	11101	0010^T
		00100	11011	0100^T
		01000 ← e	10111 ← y	1000^T
		10000	01111	1111^T
$\binom{5}{2}=10$	peso 2	00011	11100	0011^T
		00101	11010	
		00110	11001	
		01001	10110	
		⋮	⋮	

q^n sequenze

2^5 seq.

Sindrome di $y \in \mathbb{F}_q^n$: $s(y) = H \cdot y^T \in \mathbb{F}_q^{n-k}$

$$H = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$n-k=4$ righe

$n=5$ colonne

$rk(H) = n-k = 4$ ✓

$rk(H)=4$

$H \cdot G^T = 0_{((n-k)\times k)}$ ✓

$((n-k)\times n)$ $(n \times k)$

$$Hy^T = \sum_{i=1}^{n} h_i \cdot y_i$$

h_i : i-esima colonna di H $= [1\ 1\ 1\ 1]$

$\hat{x} = y - e = 10111 - 01000$

Codici di Hamming

Nel caso binario, un codice di Hamming ha $n = 2^m - 1$ (per qualche $m \geq 2$) e $k = \underbrace{2^m - 1}_{n} - m$ $(m = n - k)$; la matrice di controllo contiene <u>tutte</u> le sequenze binarie di lunghezza m non nulle.

Esempio: $m = 3$ ($\rightarrow n = 7, k = 4$; $C(7,4)$); una <u>possibile</u> matrice di controllo H è:

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$-A^T$ $\quad I_{n-k}$ $\quad$ (1) (2) (3) (4) (5) (6) (7)

$G = [I_k \mid A]$ $\qquad G$

$$G = \left[\begin{array}{cccc:ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] \qquad G = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \ \ldots\ldots$$

$$s(y) = H y^T = \sum_{i=1}^{n} \vec{h_i} \cdot y_i$$

($\vec{h_i}$: i-esima colonna di H)

$$s(y) = s(x+e) = s(x) + s(e) = s(e) = \sum_{i=1}^{n} \vec{h_i} \cdot e_i$$

(e: 0)

Distanza minima del codice è 3 $\rightarrow$ <u>corregge fino a 1 errore</u>

Possiamo assumere (ai fini della decodifica) che la config. di errore e abbia peso ≤ 1

$e_i \neq 0$ <u>al più</u> per <u>una</u> posizione

$$\rightarrow s(y) = s(e) = \boxed{\vec{h_i} \cdot e_i} \quad (i = 1, 2, \ldots, n) = \vec{h_i}$$

→ Correggi l'errore in posizione i : $\hat{x} = y - e_i$

Esempio. Seq. trasmessa : $x = 1\,0\,0\,1\,1\,0\,0$

Seq. ricevuta : $y = 1\,0\,0\,0\,1\,0\,0$

Calcolo $s(y) = Hy^T = \begin{bmatrix} 0&0&0&1&1&1&1 \\ 0&1&1&0&0&1&1 \\ 1&0&1&0&1&0&1 \end{bmatrix} \cdot \begin{bmatrix} 1\\0\\0\\0\\1\\0\\0 \end{bmatrix} = \begin{bmatrix} 0\\0\\1 \end{bmatrix} \cdot 1 + \begin{bmatrix} 1\\0\\1 \end{bmatrix} \cdot 1 = \begin{bmatrix} 1\\0\\0 \end{bmatrix} = (100)^T$

$s(y) = s(e) = 100^T \Rightarrow \vec{h_i} = \begin{bmatrix} 1\\0\\0 \end{bmatrix} \Rightarrow i = 4 \Rightarrow e = (0001000)$

$e_i = 1$, $i = 4$

$\Rightarrow \hat{e} = \vec{h_4} \Rightarrow \hat{x} = y - \hat{e} = 1000100 - 0001000 = 1001100$

→ In questo caso, la decodifica è corretta.

Decodifica per un codice di Hamming:

1. Ricevi y dal canale e calcola $s(y) = Hy^T$
2. Se $s(y) = \vec{0}$, restituisci $\hat{x} = y$
3. Altrimenti, cerca la colonna $\vec{h_i}$ di H tale che $s(y) = e_i \cdot \vec{h_i}$; restituisci $\hat{x} = y - \hat{e}$ (correggi l'i-esima posizione di y con un pezzo e_i)

Codici di Hamming generali (non necessariamente binari)

Alfabeto q-ario ($q \geq 2$); parametro $m \geq 2$

Un codice di Hamming q-ario con parametro m ha nella matrice di controllo $n = (q^m - 1)/(q-1)$ colonne, tutte <u>a coppie</u> linearmente indipendenti.

$\Rightarrow k = n - m = (q^m - 1)/(q-1) - m$

0. Poni $S = \mathbb{F}_q^m \setminus \{00\ldots0\}$

Operativamente, per costruire H:

1 - Scelgo un elemento $h_1 \in S$

2 - Elimina da S tutti i multipli di h_1:
(tutti gli elementi della forma $\alpha \cdot h_1$ con $\alpha \in \mathbb{F}_q$)

3 - Torna al punto e prosegui con $h_2, h_3, \ldots$.

$q = 3$

$0, 1, 2$

$\mathbb{F}_3$

$h_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 2 \cdot h_1$

Esempio: $\mathbb{F}_q = \mathbb{F}_3$ — $q = 3$, $m = 3$; $d_{min} = 3$.

$\rightarrow n = (q^3 - 1)/(q-1) = 13$, $k = n - m = 10$;

codice di tipo $C(13, 10, 3)$.

Matrice di controllo

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{bmatrix}$$

$$2 \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

$\frac{q^m - 1}{q - 1}$ deriva dal fatto che ci sono $q^m - 1$ possibili sequenze non-nulle di lunghezza m;

ogni sequenza ha esattamente q multipli (includendo anche se stessa)

$\rightarrow q - 1$ multipli diversi che vengono esclusi.