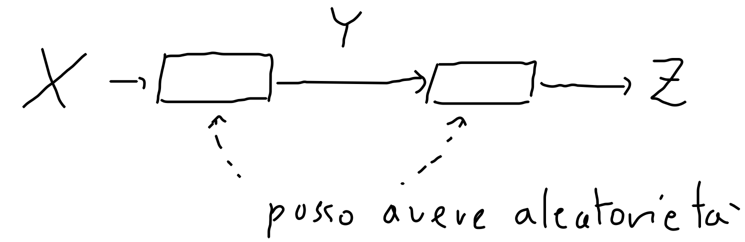
Teoremi di elaborazione dei dati

Catena di Markov:



Tre v.a. X, Y, e Z sono in catena di Markov

$(X \to Y \to Z)$ se: $\boxed{p(z|x,y) = p(z|y)}$ $\forall x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}$

data la Y, la Z non dipende dalla X

Equivalentemente: $p(x,y,z) = p(x) \cdot p(y,z|x) = p(x) \cdot p(y|x) \cdot \underbrace{p(z|x,y)}_{= p(z|y)}$

$$\to = p(x) \cdot p(y|x) \cdot p(z|y)$$

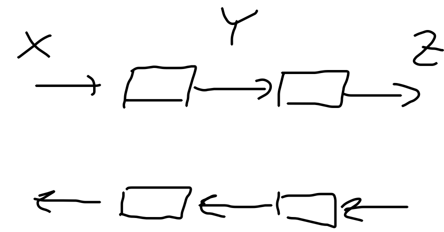
Oppure: $H(Z|X,Y) = H(Z|Y)$ $\qquad \underbrace{\mathop{E}_{(X,Y,Z)\sim p}[-\log p(Z|X,Y)]}_{H(Z|X,Y)} = \underbrace{E[-\log p(Z|Y)]}_{H(Z|Y)}$

$$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|Y)$$

Se $X \to Y \to Z$, X e Z non sono necessariamente indipendenti

però sono indipendenti condizionate ad Y:

$$I(X;Z|Y) = H(Z|Y) - H(Z|X,Y) = 0$$



⊛ Se $X \to Y \to Z$, allora vale anche $Z \to Y \to X$:

Dim. $0 = I(X;Z|Y) \overset{*}{=} H(X|Y) - H(X|Y,Z) \iff p(x|y) = p(x|y,z)$

Teo 2.7. 1° teorema di elaborazione dei dati:

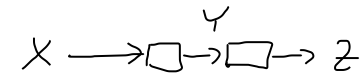
Se $X \to Y \to Z$ allora $H(X|Y) \leq H(X|Z)$ ("la Z informa di meno su X di quanto informa la Y")

Dim. $H(X|Y) \overset{*}{=} H(X|Y,Z) \leq H(X|Z)$. QED

rimuovere condizionamento aumenta l'incertezza.

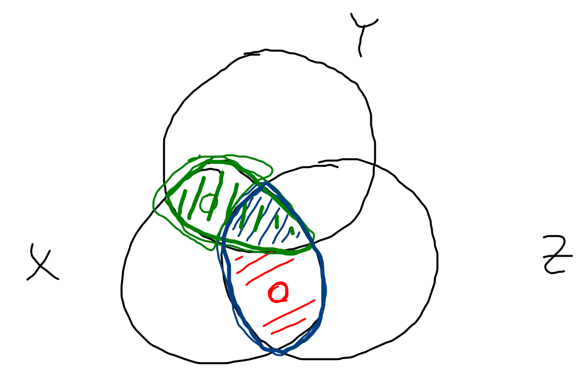
Teo 2.8 : 2° teorema di elaborazione dei dati

Se $X \to Y \to Z$ allora $I(X;Z) \leq I(X;Y)$.

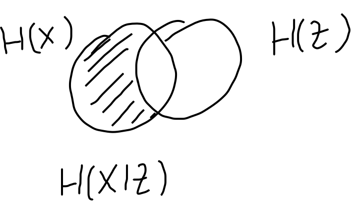


("due v.a. loutone nella catena si scambiano meno informazione di due v.a. vicine nella catena")

Dim. $I(X;Z) = H(X) - H(X|Z) \underset{\uparrow\, 1° \text{ teorema}}{\leq} H(X) - H(X|Y) = I(X;Y)$. QED.



Se $\underline{X \to Y \to Z}$, $\underbrace{I(X;Z|Y)}_{(X \cap Z) \setminus Y} = 0$



Inoltre, $I(X;Z) = I(X;Y)$

sse $I(X;Y|Z) = 0$ (equivalente a $X \to Z \to Y$)

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

$I(X;Z) \leq I(X;Y)$

Corollario : Se $Z = g(Y)$ con g funz. (deterministica) di Y, allora $I(X;Y) \overset{2° \text{ teor.}}{\geq} I(X;g(Y))$.

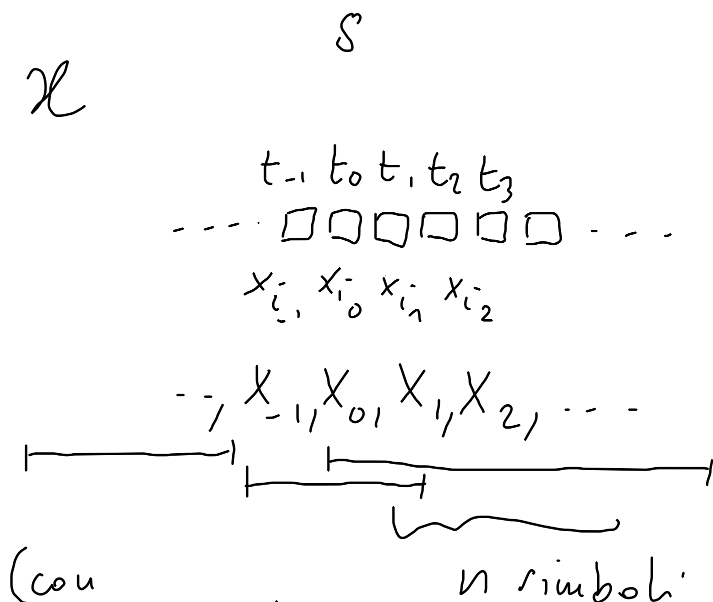
Dim. $X \to Y \to g(Y)$ sono in catena di Markov in quanto $H(g(Y)|X,Y) = 0 = H(g(Y)|Y)$. QED

Sorgente di informazione S

Ad ogni istante di tempo (discreto) t_j $(\ldots t_{-1}, t_0, t_1, \ldots)$

emette un simbolo dell'alfabeto $\mathcal{X} = \{x_1, \ldots, x_k\}$

Abbiamo una successione di v.a. $\ldots X_{-1}, X_0, X_1, X_2, \ldots$, su $\mathcal{X}$



Per descrivere la sorgente S dovremmo fornire

una famiglia infinita di d.p. del tipo

$$\Pr[(X_{j+1}, X_{j+2}, \ldots, X_{j+n}) = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})]$$

$\forall j \in \mathbb{Z}$ $\forall n \in \mathbb{N}$ $\forall x_{i_m} \in \mathcal{X}$ (con $1 \leq m \leq n$)

Servono altre ipotesi: stazionarietà e assenza di memoria

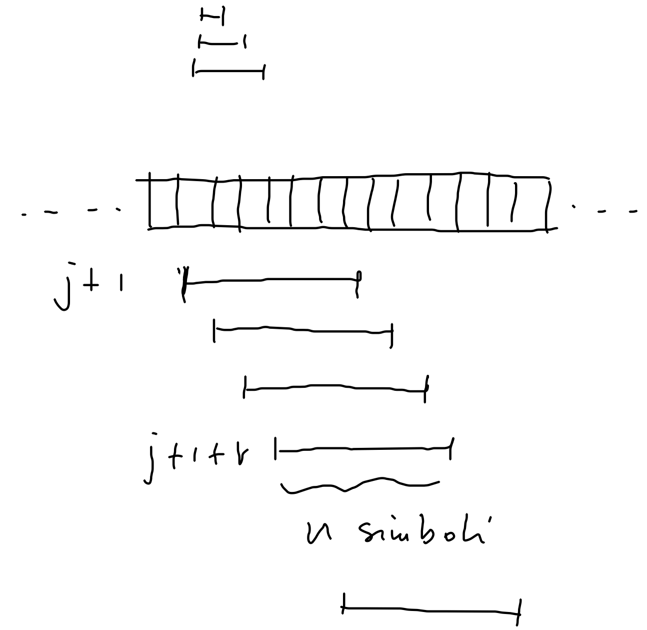
Stazionarietà:

$$\Pr[(X_{j+1}, X_{j+2}, \dots, X_{j+n}) = (x_{i_1}, x_{i_2}, \dots, x_{i_n})]$$

$$= \Pr[(X_{j+1+r}, X_{j+2+r}, \dots, X_{j+r+n}) = (x_{i_1}, x_{i_2}, \dots, x_{i_n})]$$

$$\forall r \in \mathbb{Z} \quad \forall j \in \mathbb{Z}$$

(invarianza rispetto a traslazioni nel tempo).



Assenza di memoria:

$$\Pr[(X_{j+1}, X_{j+2}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})]$$

$$= \prod_{r=1}^{n} \Pr[X_{j+r} = x_{i_r}]$$ ← sto assumendo che le $X_{j+1}, \dots, X_{j+n}$ siano indipendenti

$\boxed{X_1}\boxed{X_2}$

Lingua inglese

Carattere a caso X_1, X_2

$\Pr[X_1 = 'Q'] = 0.12\% \sim 10^{-3}$
$= 0.0012$

$\Pr[X_2 = 'U'] = 2.73\% \sim 2 \cdot 10^{-2}$
$= 0.0273$

$\Pr[X_1 = 'Q'] \cdot \Pr[X_2 = 'U']$

//

$\Pr[X_1 = 'Q', X_2 = 'U']$

(bigrammi) $\boxed{QU}$ $\sim \cancel{2 \cdot 10^{-5}}$

$\sim 10^{-3}$

Se valgono entrambe le assunzioni:

$$Pr[(X_{j+1}, X_{j+2}, \ldots, X_{j+n}) = (x_{i_1}, \ldots, x_{i_n})] = \prod_{r=1}^{n} Pr[X_r = x_{i_r}]$$

Tutte le X_r saranno identicamente distribuite

Mi basta una d.p. sull'alfabeto $\mathcal{X}$

Una d.p. $p = (p_1, \ldots, p_K)$ su $\mathcal{X} = \{x_1, \ldots, x_K\}$

Def. entropia della sorgente, $H(S)$, la quantità $H(X) = -\sum_{i=1}^{K} p_i \log p_i$.

(quantità di informazione media emessa da S).