MUTUA INFORMAZIONE CONDIZIONATA

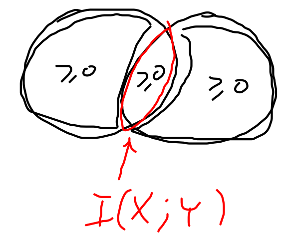
X, Y, Z v.a. $(\in \mathcal{X}, \mathcal{Y}, \mathcal{Z})$

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

$$0 \le I(X;Y|Z) \stackrel{def}{=} \mathbb{E}_{Z \sim p_Z}\left[D(p_{XY|Z} \| p_{X|Z} \cdot p_{Y|Z}) \right]$$

$$= \sum_{x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)\,p(y|z)}$$

$$= \mathbb{E}_{(X,Y,Z) \sim p_{XYZ}}\left[\log \frac{p(X,Y|Z)}{p(X|Z)\,p(Y|Z)} \right]$$



$H(X|Y)$ vs $H(X)$?

$H(X|Y) \le H(X)$

→ In particolare, se $H(X) = 0$ allora $H(X|Y) = 0$

$I(X;Y|Z)$ vs $I(X;Y)$?

Si possono definire v.a. X, Y, Z tali che:

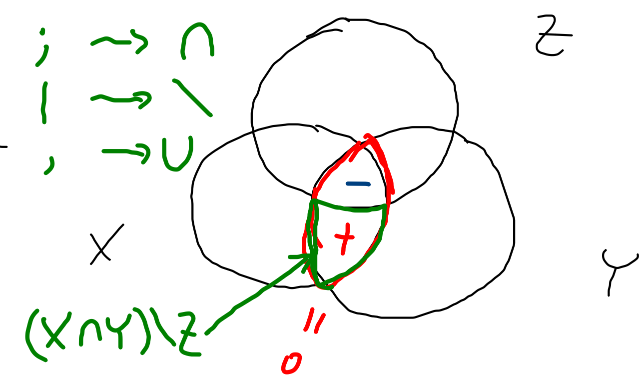
valga simultaneamente $\boxed{I(X;Y|Z) > 0 \quad e \quad I(X;Y) = 0}$ ←

Oppure tali che: $\boxed{I(X;Y|Z) = 0 \quad e \quad I(X;Y) > 0}$

$;\ \rightsquigarrow\ \cap$

$|\ \rightsquigarrow\ \setminus$

$,\ \rightsquigarrow\ \cup$



Entropia di funzioni (deterministiche) di una v.a.

Sia X una v.a. e $f : \mathcal{X} \to \mathcal{Y}$ una funzione (deterministica)

Qual è il legame tra $H(X)$ e $H(f(X))$?

Per la regola della catena,

$$H(X, f(X)) = H(X) + \cancelto{0}{H(f(X) | X)} = H(X)$$

D'altra parte, $H(X, f(X)) = H(f(X)) + H(X | f(X))$

$$\Rightarrow H(X) = H(f(X)) + \underbrace{H(X | f(X))}_{\geq 0} \geq H(f(X)) .$$

QED

(con uguaglianza se e solo se f è biettiva)

ENTROPIA DI UNA SEQUENZA DI V.A. (V.A. VETTORIALI)

sequenza di n v.a. $\longrightarrow$ $X^n = (X_1, X_2, \dots, X_n) \in \mathcal{X}^n = \underbrace{\mathcal{X} \times \mathcal{X} \times \dots \times \mathcal{X}}_{n \text{ volte}}$

$$H(X^n) = H(X_1, X_2, \dots, X_n) =$$

$$= -\sum_{\vec{x} \in \mathcal{X}^n} p(\vec{x}) \log p(\vec{x})$$

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

$$\rightarrow H(X_1, X_2) = H(X_1) + H(X_2|X_1)$$

$n=2$

Regola della catena generalizzata:

$$H(X^n) = H(X_1, \dots, X_n) = \boxed{\sum_{i=1}^{n} H(X_i | X_1, X_2, \dots, X_{i-1})}$$

$$= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_n | X_1, \dots, X_{n-1})$$

Dim. Per induzione. Banale per $n=1$. In generale, se $n>1$,

ipotesi induttiva per $n-1$

$$H(\underbrace{(X_1, \dots, X_{n-1})}, X_n) = H(X_1, \dots, X_{n-1}) + H(X_n | X_1, X_2, \dots, X_{n-1}) = \boxed{\sum_{i=1}^{n-1} H(X_i | X_1, \dots, X_{i-1}) + H(X_n | X_1, \dots, X_{n-1})}$$

regola della catena "semplice"

QED.

Corollario. $H(X^n) = H(X_1, \dots, X_n) \leq \sum_{i=1}^n H(X_i)$

Dim. $H(X^n) = \sum_{i=1}^n \underbrace{H(X_i | X_1, \dots, X_{i-1})}_{\text{Ciascuno di questi termini è al più } H(X_i)} \leq \sum_{i=1}^n H(X_i)$.

SEQUENZE DI V.A. E CONDIZIONAMENTO

Prop. $H(Y^n | X^n) = H(Y_1, \dots, Y_n | X^n) = \sum_{i=1}^n H(Y_i | X^n, Y_1, Y_2, \dots, Y_{i-1})$

→ (a) è $\leq \sum_{i=1}^n H(Y_i | X_i)$

→ (b) Inoltre, nella disug. (a) ho uguaglianza se e solo se

$$\boxed{p(\vec{y} | \vec{x}) = \prod_{i=1}^n p(y_i | x_i)}$$ per ogni n-pla $\vec{x}$ e n-pla $\vec{y}$

Dim. La parte (a) segue dalla regola della catena generalizzata:

$$H(Y^n | X^n) = \sum_i H(Y_i | \underbrace{X^n, Y_1, \dots, Y_{i-1}}_{\text{tolgo il condizionamento di tutte queste v.a. tranne } X_i}) \leq \sum_{i=1}^n H(Y_i | X_i)$$

La parte (b) si può dimostrare esplicitando le definizioni di $H(Y^n|X^n)$ e $\sum_i H(Y_i|X_i)$

Per esempio, quando $n=2$, si può verificare

$$H(Y^2|X^2) = H(Y_1, Y_2 | X_1, X_2)$$

$$= \sum_{x_1\in\mathcal{X}, x_2\in\mathcal{X}, y_1\in\mathcal{Y}, y_2\in\mathcal{Y}} \overbrace{\boxed{p(x_1,x_2,y_1,y_2)}}^{\geq 0}\Big[\overbrace{-\log p(y_1,y_2|x_1,x_2)}^{\geq 0}\Big]$$

$$\stackrel{?}{=}$$

$$\sum_{i=1}^{2} H(Y_i|X_i) = H(Y_1|X_1) + H(Y_2|X_2) =$$ (applicando la definizione e riarrangiando)

$$= \sum_{x_1,x_2\in\mathcal{X}, y_1,y_2\in\mathcal{Y}} \underbrace{\boxed{p(x_1,x_2,y_1,y_2)}}_{\geq 0}\Big[\underbrace{-\log\big(p(y_1|x_1)\cdot p(y_2|x_2)\big)}_{\geq 0}\Big]$$

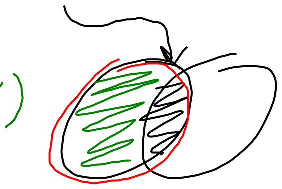
Per avere lo stesso valore, devo avere $p(y_1,y_2|x_1,x_2) = p(y_1|x_1)\cdot p(y_2|x_2)$

per ciascuna coppia (x_1,x_2) e (y_1,y_2)

Prop. (Fabris teorema 2.4): Se le v.a. $X_1, \dots, X_n$ sono tra loro indipendenti, allora

$$I(X^n; Y^n) \geq \sum_{i=1}^n I(X_i; Y_i)$$

Dim. $I(X^n; Y^n) = H(X^n) - H(X^n | Y^n)$ $\qquad I(X;Y) = H(X) - H(X|Y)$



le X_i sono indip. $\rightarrow$
$$= \sum_{i=1}^n H(X_i) - H(X^n | Y^n)$$

Per la parte (a) delle prop. precedente $\rightarrow$
$$\geq \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i | Y_i)$$

$$= \sum_{i=1}^n \big(\underbrace{H(X_i) - H(X_i | Y_i)}_{I(X_i; Y_i)} \big) = \sum_{i=1}^n I(X_i; Y_i) \quad \text{QED.}$$

Prop. (Fabris teorema 2.5)

$\rightarrow$ Se $p(\vec{y} \mid \vec{x}) = \prod_{i=1}^{n} p(y_i \mid x_i)$ (per ogni n-pla $\vec{x}$ e n-pla $\vec{y}$)

allora $I(X^n ; Y^n) \leq \sum_{i=1}^{n} I(X_i ; Y_i)$.

Dim. $I(X^n ; Y^n) = H(Y^n) - H(Y^n \mid X^n) =$

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i)$

Per la parte (b) della prop. precedentemente viste

$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i)$

$= \sum_{i=1}^{n} \big(\underbrace{H(Y_i) - H(Y_i \mid X_i)}_{I(X_i ; Y_i)} \big) = \sum_{i=1}^{n} I(X_i ; Y_i)$

QED

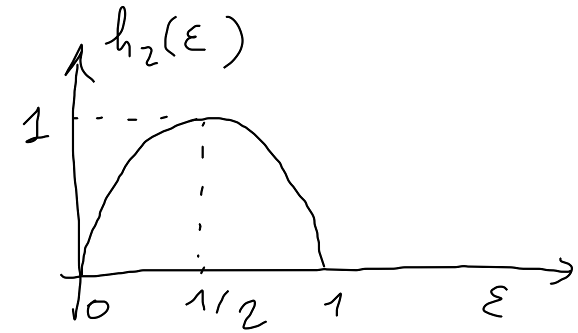
Disuguaglianza di Fano.

Siano X, $\hat{X}$ due v.a. sullo stesso alfabeto $\mathcal{X} = \{x_1, \dots, x_K\}$

(pensiamo a $\hat{X}$ come ad una stima di X)

Se $\varepsilon \overset{\text{def}}{=} \Pr[X \neq \hat{X}]$ è la prob. che la stima errata

allora: $$H(X \mid \hat{X}) \leq h_2(\varepsilon) + \varepsilon \log (K-1).$$



$$h_2(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon) \log (1-\varepsilon)$$