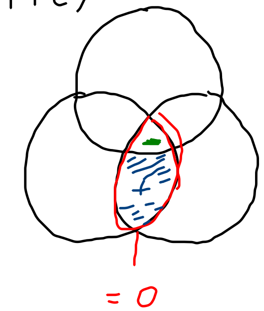
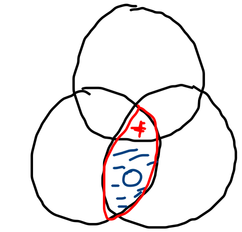
GENERALIZZAZIONI

① MUTUA INFORMAZIONE CONDIZIONATA

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



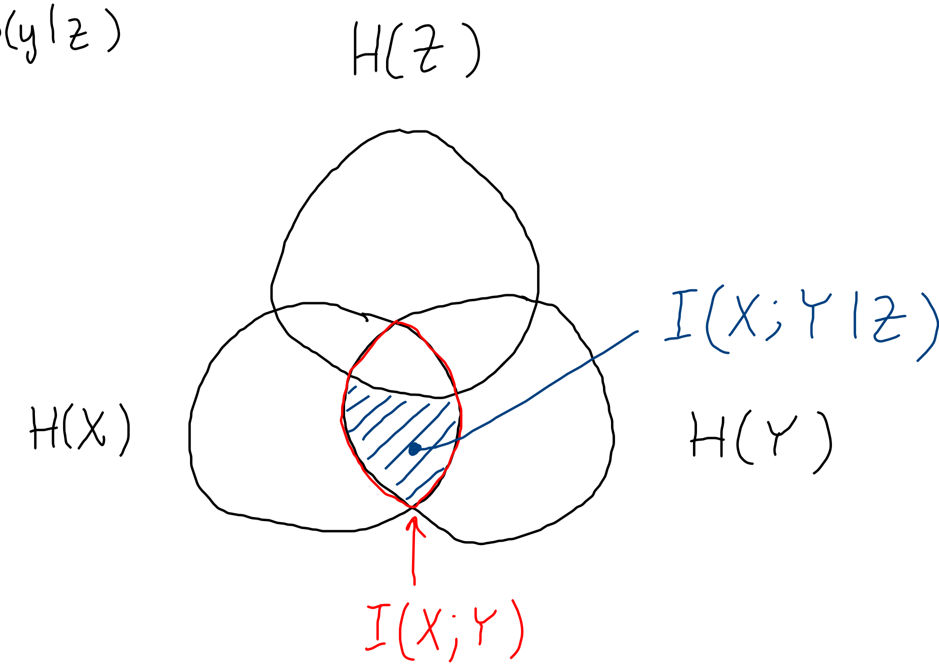
$I(X;Y)>0$
$I(X;Y|Z)=0$



$$I(X;Y|Z) \overset{def}{=} D(p_{XY|Z} \,\|\, p_{X|Z} \cdot p_{Y|Z}) \geq 0$$

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

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



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

Si può avere sia $I(X;Y)=0$ e $I(X;Y|Z)>0$

ma anche $I(X;Y)>0$ e $I(X;Y|Z)=0$

$(X \cap Y) \setminus Z$

$X;Y \mid Z$

② Condizionamento su più variabili

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

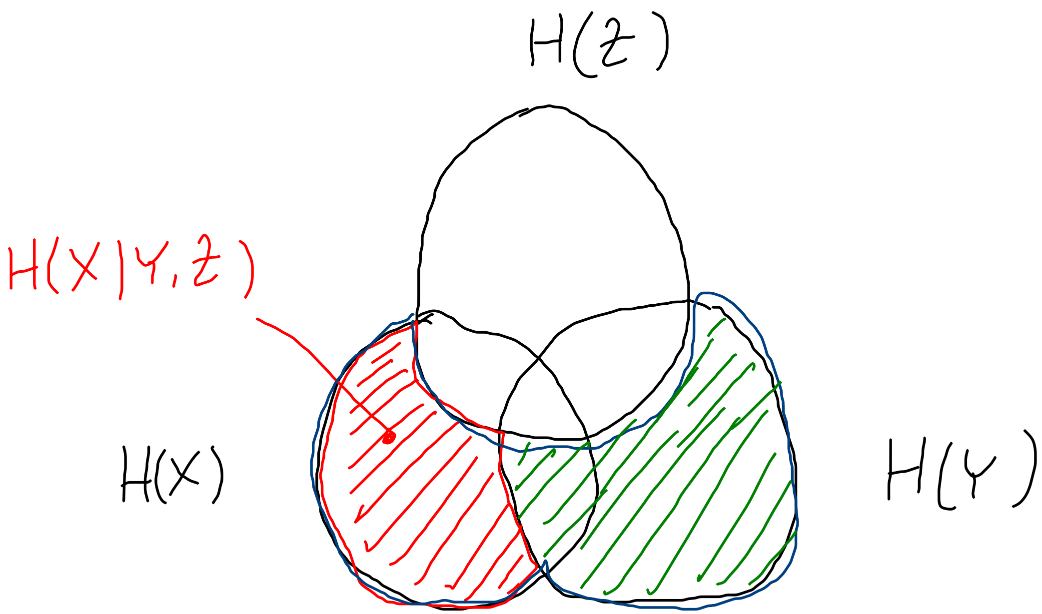
$$= \mathop{\mathbb{E}}_{(X,Y,Z)\sim p_{XYZ}}\left[-\log p(X|Y,Z)\right] \geq 0$$

Regole delle catene

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

$H(X,Y|Z)$

$H(X,Y|Z) = \underline{H(Y|Z)} + \underline{H(X|Y,Z)}$



$X|Y,Z$ $\qquad$ $X,Y|Z$

$X\setminus(Y\cup Z)$ $\qquad$ $(X\cup Y)\setminus Z$

③ ENTROPIA DI UN VETTORE

$$X^n = (X_1, X_2, \dots, X_n) \in \mathcal{X}^n \quad \text{(insieme prodotto cartesiano } \overbrace{\mathcal{X} \times \mathcal{X} \times \dots \times \mathcal{X}}^{n \text{ volte}} \text{)}$$

$H((X,Y))$

$$H(X^n) = H(X_1, X_2, \dots, X_n) = -\sum_{\vec{x} \in \mathcal{X}^n} p(\vec{x}) \log p(\vec{x})$$

REGOLA DELLA CATENA (GENERALIZZATA)

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

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

Dim. Per induzione. Ovvia per $n=1$. Assumiamo sia vera per $n-1$.

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

$$\overset{\text{ipot. induttiva}}{=} \sum_{i=1}^{n-1} H(X_i | X_1, \dots, X_{i-1}) + H(X_n | X_1, \dots, X_{n-1}) = \sum_{i=1}^{n} H(X_i | X_1, \dots, X_{i-1}).$$

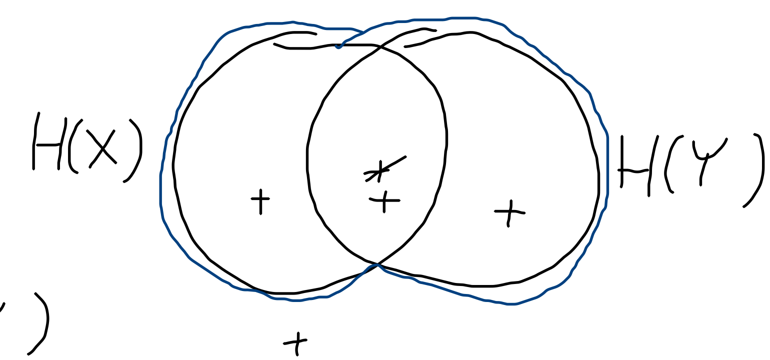
QED

$$H(X^n) = H(X_1, \dots, X_n) \overset{\text{catene (gen.)}}{=} \sum_{i=1}^{n} \underbrace{H(X_i \mid X_1, \dots, X_{i-1})}_{\le H(X_i)} \le \sum_{i=1}^{n} H(X_i)$$

$$H(X,Y) \gtrless H(X) + H(Y)$$

$$\boxed{H(X,Y) \le H(X) + H(Y)}$$

$$H(X,Y) = H(X) + H(Y) - \underbrace{I(X;Y)}_{\ge 0} \le H(X) + H(Y)$$



$\mathcal{Y} = \{y_1, \dots, y_k\}$

$n = 2$ $\quad \mathcal{Y} = \{y_1, y_2\}$ $\quad \mathcal{X} = \{x_1, x_2\}$

$p(y_1, y_1 \mid x_1, x_1) = \dots$

$p(y_1, y_2 \mid x_1, x_1) = \dots$

Prop. (a) $H(Y^n \mid X^n) = H(Y_1, \dots, Y_n \mid X^n) \overset{\text{catene gen.}}{=} \sum_{i=1}^n H(Y_i \mid X^n, Y_1, \dots, Y_{i-1})$

è $\leq \sum_{i=1}^n H(Y_i \mid X_i)$

(b) ed è $=$ sse $p(\vec{y} \mid \vec{x}) = p(y_{i_1}, y_{i_2}, \dots, y_{i_n} \mid x_{j_1}, x_{j_2}, \dots, x_{j_n}) = \prod_{r=1}^n p(y_{i_r} \mid x_{j_r})$

$\boxed{\forall \vec{y} \in \mathcal{Y}^n, \forall x \in \mathcal{X}^n}$

Dim. (a) segue dalla regola della catena generalizzata

poiché $H(Y_i \mid X^n, Y_1, \dots, Y_{i-1}) \leq H(Y_i \mid X_i)$.

Rimuovere il condizionamento può solo aumentare l'entropia

(b) Si può dimostrare esplicitando le definizioni

Per esempio, per $n = 2$, esplicitando $H(Y_1, Y_2 \mid X_1, X_2) = -\sum_{x_1, x_2 \in \mathcal{X}, y_1, y_2 \in \mathcal{Y}} \overbrace{p(x_1, x_2, y_1, y_2)}^{\geq 0} \log \boxed{p(y_1, y_2 \mid x_1, x_2)}$

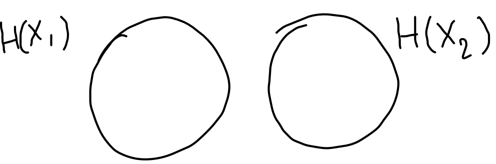
e $H(Y_1 \mid X_1) + H(Y_2 \mid X_2) = -\sum_{x_1, x_2 \in \mathcal{X}, y_1, y_2 \in \mathcal{Y}} \underbrace{p(x_1, x_2, y_1, y_2)}_{\geq 0} \log \boxed{(p(y_1 \mid x_1) \cdot p(y_2 \mid x_2))}$

<u>Prop</u>. (Teo 2.4) Se le X_i sono indipendenti, allora $I(X^n;Y^n) \geq \sum_{i=1}^n I(X_i;Y_i)$.

Dim. $I(X^n;Y^n) = H(X^n) - H(X^n|Y^n) = \sum_{i=1}^n H(X_i) - \underbrace{H(X^n|Y^n)}_{\leq \sum_{i=1}^n H(X_i|Y_i)}$

Usando le prop. precedente,

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



$I(X_1;X_2) = 0$

$I(X;Y) = H(X) - H(X|Y)$

Prop. (Teo 2.5) Se $p(\vec{y}|\vec{x}) = p(y_{i_1}, \dots, y_{i_n} | x_{j_1}, \dots, x_{j_n}) = \prod_{r=1}^{n} p(y_{i_r} | x_{j_r})$

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

<u>Dim.</u> $I(X^n; Y^n) = H(Y^n) - H(Y^n | X^n) \overset{\text{prop. parte (b)}}{=} H(Y^n) - \sum_{i=1}^n H(Y_i | X_i)$

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

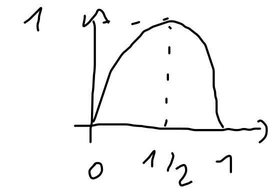
$I(X;Y) = H(Y) - H(Y|X)$

Disuguaglianza di Fano. Siano X e $\hat{X}$ due v.a. su $\mathcal{X} = \{x_1, \dots, x_K\}$

(Consideriamo uno scenario in cui $\hat{X}$ è una stima di X)

Chiamiamo $\varepsilon \overset{def}{=} \Pr[X \neq \hat{X}]$ (prob. che la stima sia errata).

Possiamo limitare l'incertezza su X una volta che la stima $\hat{X}$ sia nota?



(Teo 2.6) Si ha $H(X \mid \hat{X}) \le h_2(\varepsilon) + \varepsilon \underbrace{\log(K-1)}_{\ge 0}$.

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

Dim Definiamo $Z = \begin{cases} 1 & \text{se } X \neq \hat{X} \quad (\text{prob. } \varepsilon = \Pr[Z=1]) \\ 0 & \text{se } X = \hat{X} \quad (\text{prob } 1-\varepsilon = \Pr[Z=0]) \end{cases}$

Per la regola della catena,

$$H(X,Z \mid \hat{X}) = H(X \mid \hat{X}) + \underbrace{\cancel{H(Z \mid X, \hat{X})}}_{0}$$

0 (se conosco sia X che $\hat{X}$, conosco Z)

ma anche $= H(Z \mid \hat{X}) + H(X \mid \hat{X}, Z)$

$$\overbrace{\Pr[Z=1]}^{\varepsilon} H(X \mid Z=1, \hat{X}) + \underbrace{\Pr[Z=0]}_{(1-\varepsilon)} \underbrace{H(X \mid Z=0, \hat{X})}_{=0}$$

$$H(X \mid \hat{X}) = H(Z \mid \hat{X}) + H(X \mid \hat{X}, Z) \le \underbrace{H(Z)}_{= h_2(\varepsilon)} + H(X \mid \hat{X}, Z) \le h_2(\varepsilon) + \varepsilon \cdot \underbrace{\log(K-1)}_{\text{entropia massima quando varia su } K-1 \text{ simboli}}$$

. QED.