Entropia di Shannon / 22-04

Data X var casuale che assume valori $x_1, \dots, x_N$ (discreti)

$0 \leq p(x_1), \dots, p(x_N) \leq 1$

$\sum p = 1$

$V_x = \{x_1, \dots, x_N\}$

$|V_x| = N$ cardinalità

$$H(X) = \sum_{i=1}^{N} p(x_i) \log_2 \frac{1}{p(x_i)}$$

• $p = \frac{1}{2}$ $\quad x_1 = T \quad x_2 = C \quad p(x_1) = p(x_2) = \frac{1}{2}$

$H(X) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1$ ottengo un bit di informazione

equiprobabilità = max incertezza

• $p(T) = \frac{7}{8} \quad p(C) = \frac{1}{8}$

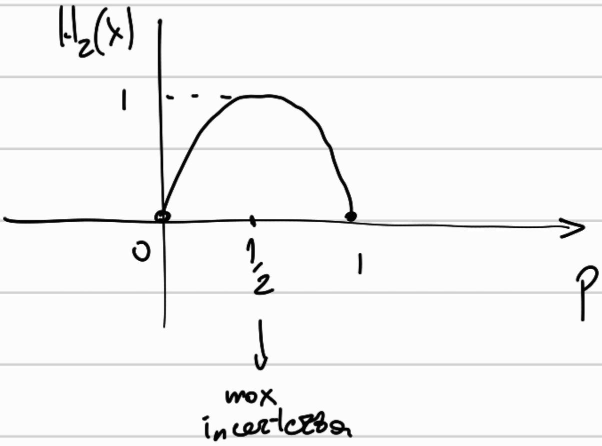
$H(X) \simeq \frac{7}{8} \cdot 0{,}19 + \frac{1}{8} \cdot 3 \simeq 0{,}6$ sicuramente più piccolo di 1

In una moneta onesta ho molte aspettative (valore atteso alto) mentre in una truccata ne ho meno ($\frac{7}{8}$ mi dà più certezza che accade T) (si abbassa il valore atteso)

$$\underset{\substack{\downarrow \\ \text{ho 2 valori di } X}}{H_2}(X) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

e se $1-p = 0$?

studiando gli infiniti x cresce "più velocemente" di log



$(0 - \ln p) = (\ln 1 - \ln p) = \ln \frac{1}{p}$

$- p \overbrace{\ln p}^{''} - (1-p) \ln (1-p)$ riscrittura

se derivo per cercare il max $\quad - \ln p \cancel{-1} + \ln(1-p) \cancel{+1} = 0$

$\ln(1-p) = \ln(p)$

$X.\ x_1 \dots x_N \quad p_i = p(x_i) = \frac{1}{N} \quad \forall i$ cioè equiprobabile

$$H(X) = \sum_i p_i \log_2 \frac{1}{p_i}$$

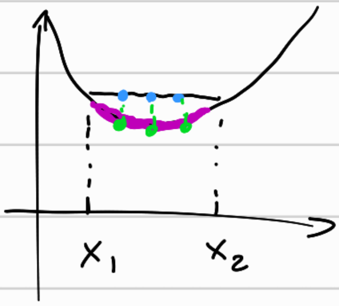
$$= \sum_i \frac{1}{N} \cdot \log_2 N = N \cdot \frac{1}{N} \log_2 N = \log_2 N$$

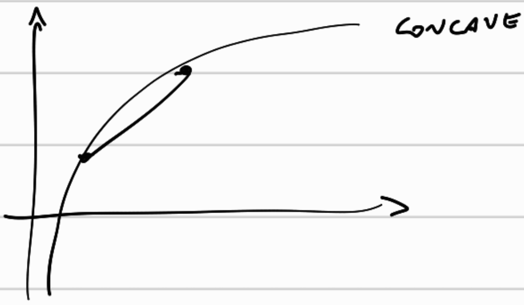
sommare N volte equivale a moltiplicare per N

se ora p_i generiche (non equiprobabile dunque rispetto a N valori)

$H(X) \leq \log_2 N$ come successo per la moneta di prima

$\leq$ equiprobabile





$$x = t x_2 + (1-t) x_1$$

f concava se $\geq$

$$0 \leq t \leq 1$$

$t=0$ ho x_1
$t=1$ ho x_2

mon mano che cresce t descrivo un punto fra x_1 e x_2

f e' convessa se $f(t x_2 + (1-t) x_1) \boxed{\leq} t f(x_2) + (1-t) f(x_1)$

calcolata in quel punto x

generalizzazione

$$E[f(x)] \geq f(E[X])$$

$$H(X) = \sum p_i \log 1/p_i$$

$$\leq \log_2 \sum_i p_i \, 1/p_i = \log_2 N$$

perché concava il valore di f e' maggiore

$p(1) = p(2) = p(3) = p(4) = 1/16$
$p(5) = p(6) = 1/8$
$p(7) = p(8) = 1/4$

dado a 8 facce, calcolare l'entropia:

$$4 \cdot \frac{1}{16} 4 + 2 \frac{1}{8} 3 + 2 \frac{1}{4} \cdot 2 = 1 + \frac{6}{8} + 1 = 2{,}75 < 3$$

$\downarrow$

ottenuto da equiprobabilità: $1/8$

$\sum_i^8 \frac{1}{8} \cdot \log_2 \frac{1}{1/8} = 8 \cdot \frac{1}{8} \cdot 3 = 3$

(valore atteso max)

Entropia congiunta

Dati X, Y discrete $x_1 \dots x_N$ $\quad y_1 \dots y_M$

$0 \le p(x_i, y_j) \le 1 \qquad \sum_{i=1}^{N} \sum_{j=1}^{M} p(x_i, y_j) = 1$

$$H(X,Y) = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{1}{p(x_i, y_j)}$$

Entropia condizionata

$$H(X \mid Y = \underset{\text{FISSATA}}{y_j}) = \sum_i p(x_i \mid y_j) \log_2 \frac{1}{p(x_i, y_j)}$$

$$H(X \mid Y) = \sum_j p_Y(y_j) \Big(\underbrace{\sum_i p(x_i \mid y_j) \log_2 \frac{1}{p(x_i, y_j)}}_{\text{si verifica per ogni } Y} \Big)$$

X faccia del dado
Y pari / dispari faccia

$P(X,Y) \neq P_X(x)\, P_Y(Y)$ non sono indipendenti

$p(1,p) = p(3,p) = p(5,p) = \emptyset$
$p(2,p) = p(4,p) = p(6,p) = 1/6$ p che esce 2 e sia pari

$p(1,d) = p(3,d) = p(5,d) = 1/6$
$p(2,d) = p(4,d) = p(6,d) = \emptyset$

$P(X,Y) = p(x)\, p(Y|X) = \overset{\text{p pari o dispari}}{p(Y)}\, p(X|Y)$

$p(3,p) = 1/6 \cdot \underset{\substack{\downarrow \\ \text{p che sia pari} \\ \text{sapendo sia uscito 3}}}{\emptyset} = \emptyset$

$p(3,p) = 1/2 \cdot \underset{\substack{\text{p che sia 3} \\ \text{sapendo che sia} \\ \text{pari}}}{\emptyset}$

$p(3) = p(3,p) + p(3,d) = 1/6$
marginale

$p(\text{pari}) = p(1,p) + p(2,p) + \ldots + p(6,p) = 1/2$

$p(3,p) = \emptyset \neq 1/6 \cdot 1/2$ infatti sono dipendenti

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

Dimostrazione

$$H(X,Y) = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{1}{p(x_i, y_j)}$$

$$= \sum_i \sum_j p_X(x_i)\, p(y_j|x_i) \left[\log_2 \frac{1}{p_X(x_i)} + \log_2 \frac{1}{p(y_j|x_i)} \right]$$

tolgo ciò che non dipende da i e j

$$= \sum_i p_X(x_i) \log_2 \frac{1}{p_X(x_i)} \underbrace{\sum_j p(y_j|x_i)} + \text{''} \cdot$$

↳ medesimi calcoli

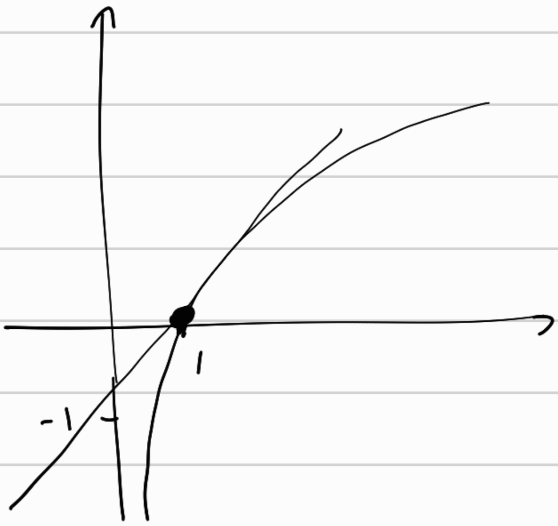
sto sommando tutte le prob di y_j (dato x_i), la somma è 1 (es. somma sia pari dopo essere uscito un num fra 1 e 6 è uno)

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

Ma allora $H(X|Y) \leq H(X)$ disuguaglianza fondamentale

↓

sapere qualcosa che è successo riduce l'entropia (riduce l'aspettativa, il valore atteso)



$$\ln t \leq t - 1$$

$$\log_2 t \leq (t-1) \log_2 e$$

↓

$$\ln t = \frac{\log_2 t}{\log_2 e}$$

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

$H(X,Y) \leq H(Y) + H(X)$ perché $H(X|Y) \leq H(X)$

(la condizionata ha meno aspettazione)

$$H(X|Y) - H(X) \leq 0$$

$$= \sum_j p_Y(y_j) \sum_i p(x_i|y_j) \log_2 \frac{1}{p(x_i|y_j)} - \sum_i p_X(x_i) \log_2 \frac{1}{p_X(x_i)}$$

$$= \sum_i \sum_j p(x_i, y_j) \log_2 \frac{1}{p(x_i|y_j)} - \sum_i \sum_j p(x_i, y_j) \log_2 \frac{1}{p_X(x_i)}$$

$$\left[p_X(x_i) = \sum_j p(x_i | y_j) \right]$$

marg. di X ciclo le y

$$= \underbrace{\sum_i \sum_j p(x_i\, y_j)}_{=1} \left[\log_2 \frac{p_X(x_i)}{p(x_i|y_j)} \right] \leq \sum_i \sum_j p(x_i, y_j) \left[\frac{p_X(x_i)}{p(x_i|y_j)} - 1 \right]$$

per l'identità di prima $\log_2 t \leq (1-t) \log_2 e$

$$= \sum_i \sum_j p_Y(y_j) \frac{\cancel{p(x_i|y_j)}}{\cancel{p(x_i|y_j)}} p_X(x_i) = 1 (\cdot \log_2 e)$$

fuori dalle sommatorie

dunque con il passo sopra: $(1-1) \log_2 e = 0$