Entropia

Informazione associata a valore x avente probabilitá $p(x)$ é

$$i(x) = \log_2 \frac{1}{p(x)}$$

Nota: Eventi meno probabili danno maggiore informazione

Entropia di v.c. $X \sim P$: informazione media elementi di $\mathcal{X}$

$$H(X) = \sum_{x \in \mathcal{X}} p(x) i(x) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$$

Rappresenta l'incertezza su esito X

Proprietà dell'entropia

- L'unitá di misura dell'entropia sono i bits
  (Usiamo $\log$ in base 2)
- Se si cambia la base del logaritmo, il valore dell'entropia cambia solo di un fattore costante

Lemma $H_b(X) = \frac{H(X)}{\lg b}$

Dim.

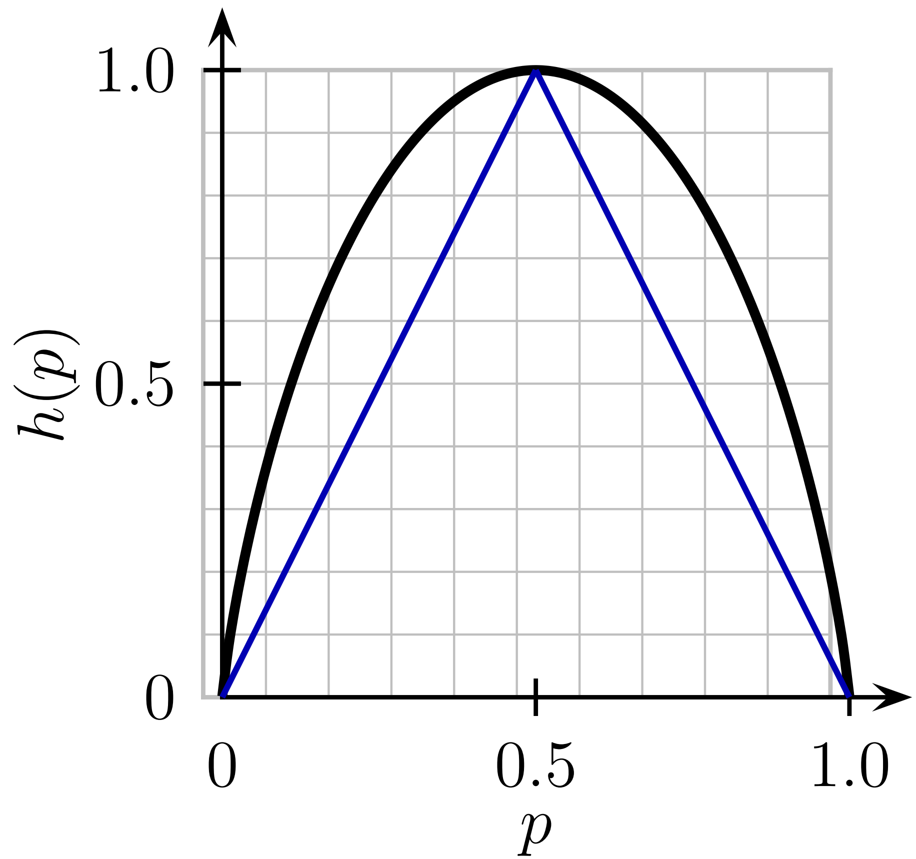
$$H_b(X) = \sum_{x \in \mathcal{X}} p(x) \lg_b \frac{1}{p(x)} = \sum_{x \in \mathcal{X}} p(x) \frac{\lg \frac{1}{p(x)}}{\lg b}$$

$$= \frac{1}{\lg b} \sum_{x \in \mathcal{X}} p(x) \lg \frac{1}{p(x)} = \frac{H(X)}{\lg b}$$

□

Entropia Binaria (v.c. di Bernulli)

$$X = \begin{pmatrix} a & b \\ p & 1-p \end{pmatrix} \qquad H(X) = p \lg \frac{1}{p} + (1-p) \lg \frac{1}{(1-p)} \stackrel{def}{=} h(p)$$



$$h'(p) = \lg \frac{1-p}{p}$$
$$h''(p) = -\frac{\log e}{p(1-p)}$$
$$h(p) \geq 2 \min\{p, 1-p\}$$

Entropia congiunta

Def. Date due v.c. X e Y con d.d.p. congiunta $p(x, y)$, definiamo *entropia congiunta* la quantità :

$$H(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \lg \frac{1}{p(x, y)} = E\left[\lg \frac{1}{p(X, Y)}\right]$$

$p(x,y)$	$y = 0$	$y = 1$
$x = 0$	$1/4$	$1/4$
$x = 1$	$1/2$	0

$$\begin{aligned} H(X,Y) &= H(1/4, 1/4, 1/2) \\ &= 2\frac{1}{4}\log 4 + \frac{1}{2}\log 2 = \frac{3}{2} \text{ bits} \end{aligned}$$

Entropia condizionata

Entropia condizionata $H(Y/X)$:

$$H\left(Y/X\right) = \sum_{x\in\mathcal{X}} p\left(x\right) H\left(Y/X=x\right)$$

$H\left(Y/X=x\right)$: entropia di Y sapendo che $X=x$

$p(x,y)$	$y=0$	$y=1$	$p(x)$	$H(Y/X=x)$
$x=0$	$1/2$	$1/4$	$3/4$	$H(2/3,1/3)=h(1/3)$
$x=1$	$1/4$	0	$1/4$	$H(1,0)=h(1)=0$

$H(Y/X) = \frac{3}{4}h\left(\frac{1}{3}\right)+\frac{1}{4}h\left(1\right) = \frac{3}{4}h\left(\frac{1}{3}\right)$ bits

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

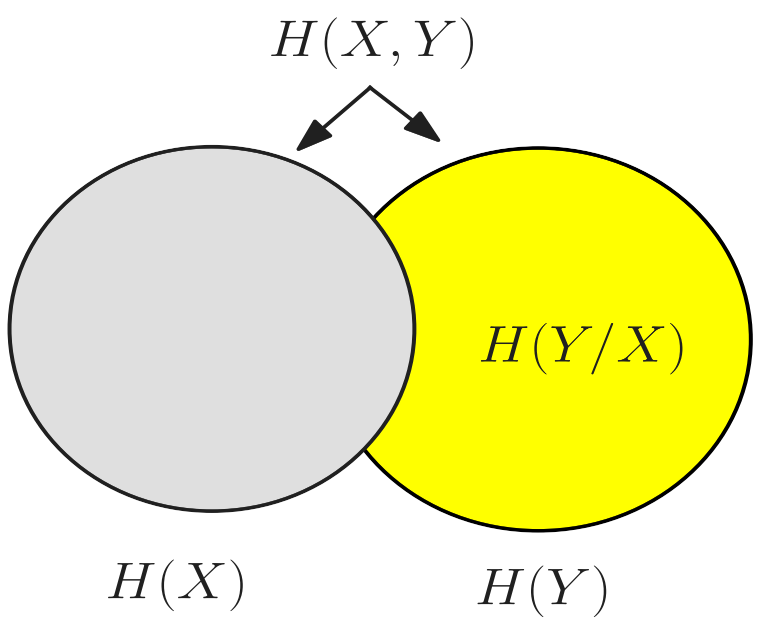
Entropia condizionata

$$
\begin{aligned}
H\left(Y/X\right) &= \sum_{x\in\mathcal{X}} p\left(x\right) H\left(Y/X = x\right) \\
&= \sum_{x\in\mathcal{X}} p\left(x\right) \sum_{y\in\mathcal{Y}} p\left(y/x\right) \lg \frac{1}{p\left(y/x\right)} \\
&= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} p\left(x\right) p\left(y/x\right) \lg \frac{1}{p\left(y/x\right)} \\
&= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} p\left(x, y\right) \lg \frac{1}{p\left(y/x\right)} = E\left[\lg \frac{1}{p(y/x)}\right]
\end{aligned}
$$

Entropia condizionata

$$H\left(Y/X\right) = \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} p\left(x,y\right)\lg\frac{1}{p\left(y/x\right)} = E\left[\lg\frac{1}{p(y/x)}\right]$$

rappresenta l'informazione addizionale media di Y nota X.



Regola della catena

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

$$\begin{aligned}
H(X,Y) &= \sum_{x,y} p(x,y) \lg \frac{1}{p(x,y)} = \sum_{x,y} p(x,y) \lg \frac{1}{p(y/x)\, p(x)} \\
&= \sum_{x,y} p(x,y) \left[\lg \frac{1}{p(x)} + \lg \frac{1}{p(y/x)} \right] \\
&= \sum_{x,y} p(x,y) \lg \frac{1}{p(x)} + \sum_{x,y} p(x,y) \lg \frac{1}{p(y/x)} \\
&= \sum_{x} \left(\sum_{y} p(x,y) \right) \lg \frac{1}{p(x)} + H(Y/X) \\
&= \sum_{x} p(x) \lg \frac{1}{p(x)} + H(Y/X) = H(X) + H(Y/X)
\end{aligned}$$

Regola della catena: dimostrazione alternativa

Consideriamo la sequente dimostrazione alternativa:

$$
\begin{aligned}
H(X,Y) &= E\left[\lg \frac{1}{p(X,Y)}\right] = E\left[\lg \frac{1}{p(Y/X)p(X)}\right] \\
&= E\left[\lg \frac{1}{p(X)} + \lg \frac{1}{p(Y/X)}\right] \\
&= E\left[\lg \frac{1}{p(X)}\right] + E\left[\lg \frac{1}{p(Y/X)}\right] \\
&= H(X) + H(Y/X)
\end{aligned}
$$

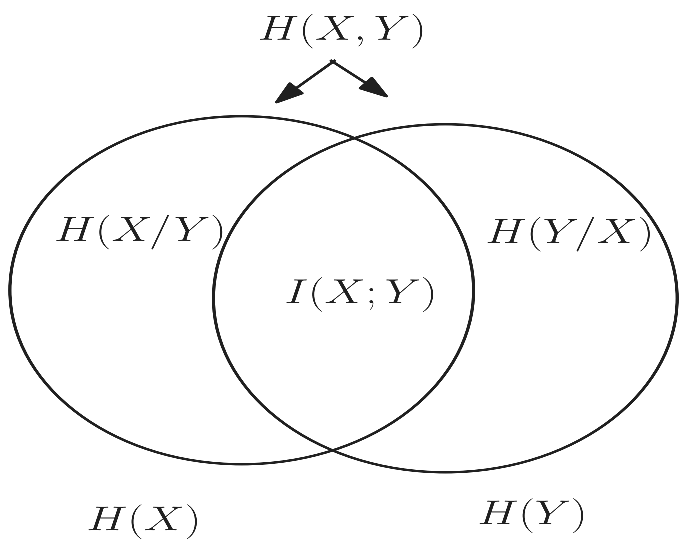
Il $\log$ trasforma prodotti di probabilitá in somme di entropie

Mutua Informazione

Date due v.c. X e Y con d.p. congiunta $p(x, y)$, la *mutua informazione* è definita come

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

La mutua informazione rappresenta i bit di informazione che una delle variabili fornisce circa l'altra.



Nota: uso ";" (es $I(X,Z;Y) \neq I(X;Z,Y)$)

Mutua Informazione

Es. Lanciamo 10 monete:
X rappresenta i valori delle prime 7 monete,
Y quelli delle ultime 5.

$H(X) = 7 \qquad H(Y) = 5, \qquad H(X/Y) = 5 \quad H(Y/X) = 3$

$\Rightarrow \quad I(X;Y) = I(Y;X) = 2$

Mutua Informazione

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

$$\begin{aligned} I(X;Y) - I(Y;X) &= H(X) - H(X/Y) - H(Y) + H(Y/X) \\ &= H(X) + H(Y/X) - (H(Y) + H(X/Y)) \\ &= H(X,Y) - H(X,Y) = 0 \end{aligned}$$

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

Mutua Informazione Condizionata

$I(X;Y/Z) = H(X/Z) - H(X/YZ) = H(Y/Z) - H(Y/XZ)$

Nota: Il condizionamento di Z si applica SIA ad X che ad Y

$$\begin{aligned} I(X;Y/Z) &= H(X/Z) - H(X/YZ) \\ &= H(X/Z) - (H(X,Y/Z) - H(Y/Z)) \\ &= H(X/Z) + H(Y/Z) - H(X,Y/Z) \end{aligned}$$

Regola della catena per la mutua informazione

$$I(X_1, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y/X_1, \ldots, X_{i-1})$$

Nota:

$I(X_1, \ldots, X_n; Y) = H(X_1, \ldots, X_n) + H(Y) - H(X_1, \ldots, X_n, Y)$

Riepilogo/Anteprima

N.B.: Le disugualianze saranno provate in seguito

Entropia $H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}$

- $H(X) \geq 0$, uguaglianza sse $\exists\ x$ t.c. $p(x) = 1$
- $H(X) \leq \log |\mathcal{X}|$, uguaglianza sse $p(x) = 1/|\mathcal{X}|\ \forall x \in \mathcal{X}$
- Il condizionamento non aumenta l'entropia
  $H(X/Y) \leq H(X)$ uguaglianza sse X, Y indipendenti
- Regola della catena:
  $H(X, Y) = H(X) + H(Y/X) \leq H(X) + H(Y)$,
  uguaglianza sse X, Y indipendenti
  $H(X_1, \dots, X_n) = \sum_{i=1}^{n} H(X_i/X_{i-1} \dots x_1) \leq \sum_{i=1}^{n} H(X_i)$,
  uguaglianza sse $X_1, \dots, X_n$ indipendenti

Riepilogo/Anteprima

Mutua Informazione:
$I(X;Y) = H(X) - H(X/Y) = H(X) + H(Y) - H(X,Y)$

- Simmetrica e positiva: $I(X;Y) = I(Y;X) \geq 0$, uguaglianza sse X, Y indipendenti

  Infatti

  $$I(X;Y) = H(X) - H(X/Y) \geq H(X) - H(X) = 0$$

  con $H(X/Y) = H(X)$ sse X, Y indipendenti

Funzioni concave/convesse

Def. Una funzione $f(x)$ si dice *concava su un intervallo* (a, b) *se* $\forall x_1, x_2 \in (a, b)$ *e* $\forall 0 \leq \lambda \leq 1$

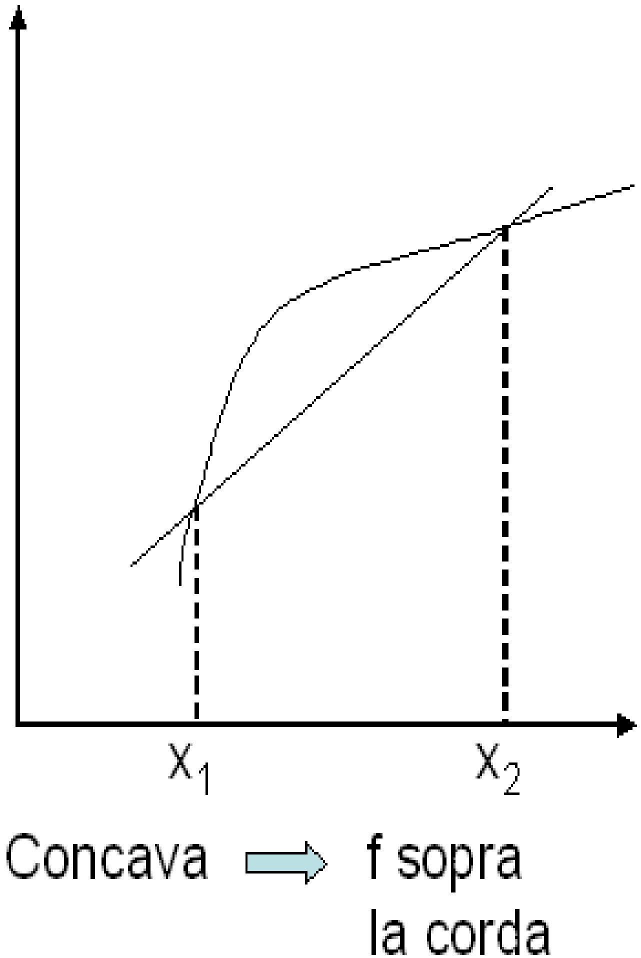
$$f(\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \geq \lambda \cdot f(x_1) + (1 - \lambda) \cdot f(x_2)$$

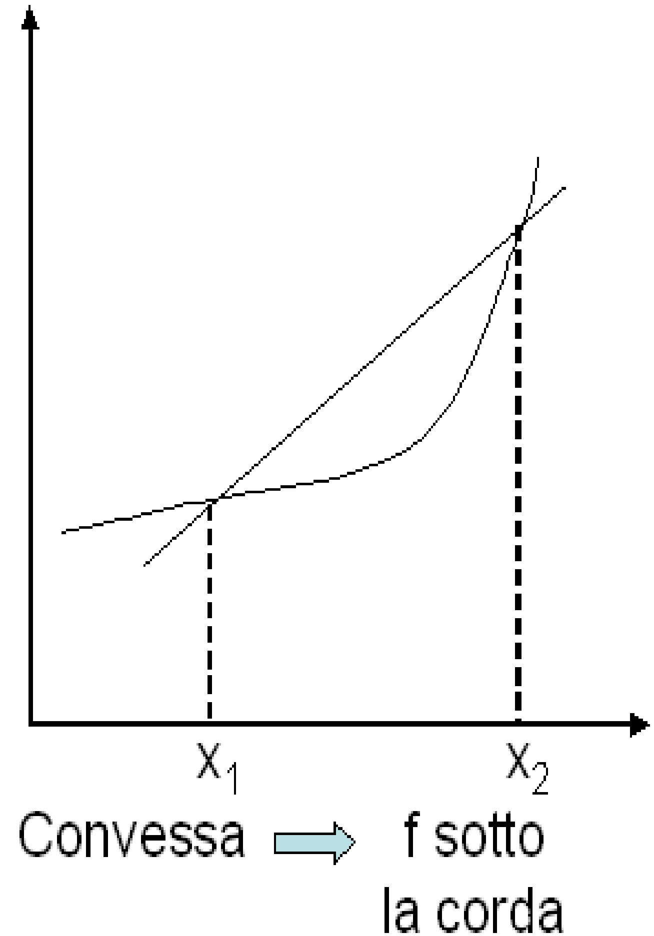
f è strettamente concava se la disuguaglianza é stretta per $0 < \lambda < 1$.

Esempio $\lg x$ è una funzione strettamente concava di x.

Def. Una funzione $f(x)$ si dice *convessa su un intervallo* (a, b) *se* $-f(x)$ *è concava sull' intervallo* (a, b).

Esempio $x^2, x \lg x$ *sono funzioni strett. convesse di* x.


x_1
x_2
Concava
f sopra
la corda


x_1
x_2
Convessa
f sotto
la corda

Diseguaglianza di Jensen

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \qquad f\text{: funzione}$$

$$f\left(X\right) = \begin{pmatrix} f(x_1) & f(x_2) & \dots & f(x_n) \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

- Risulta $E[f(X)] \leq f(E[X])$,
- Se f strettamente concava: $E[f(X)] = f(E[X])$, sse X è concentrata in un' unico punto (cioé é costante)

La dimostrazione procede per induzione su $|\mathcal{X}|$.

Base induzione: $|\mathcal{X}| = 2$.

Si consideri la v.c.

$$X = \begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix}$$

Per la definizione di funzione concava, si ha che

$$E\left[f\left(X\right)\right] = p_1 f(x_1) + p_2 f(x_2) \leq f(p_1 x_1 + p_2 x_2) = f\left(E\left[X\right]\right).$$

Se f strett. concava l'uguaglianza vale sse $p_1 = 1$ oppure $p_2 = 1$.

Passo induttivo: Supponiamo che la disuguaglianza di Jensen sia verificata per $|\mathcal{X}| = k-1$. Dimostriamo che la disuguaglianza è verificata per $|\mathcal{X}| = k$.

$$
\begin{aligned}
E\left[f\left(X\right)\right] &= \sum_{i=1}^{k} p_i f\left(x_i\right) = p_k f\left(x_k\right) + \sum_{i=1}^{k-1} p_i f(x_i) \\
&= p_k f(x_k) + (1-p_k)\left(\sum_{i=1}^{k-1} \frac{p_i}{(1-p_k)} f(x_i)\right)
\end{aligned}
$$

Osserviamo che $\sum_{i=1}^{k-1} \frac{p_i}{(1-p_k)} f(x_i) = E[f(X')]$ dove X' è la v.c.

$$X' = \begin{pmatrix} x_1 & \dots & x_{k-1} \\ \frac{p_1}{1-p_k} & \dots & \frac{p_{k-1}}{1-p_k} \end{pmatrix}$$

L'ipotesi induttiva implica $E[f(X')] \leq f(E[X'])$ per cui risulta:

$$E[f(X)] \leq p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} \frac{p_i}{1-p_k} x_i\right).$$

Applichiamo la definizione di funzione concava al termine destro

Otteniamo

$$\begin{aligned} E[f(X)] &\leq f\left(p_k x_k + (1-p_k)\sum_{i=1}^{k-1}\frac{p_i}{1-p_k}x_i\right) \\ &= f\left(\sum_{i=1}^{k} p_i x_i\right) = f\left(E\left[X\right]\right). \end{aligned}$$

Se f strettamente concava: uguaglianza sse tutti $\leq$ sono $=$

- ($=$ in def. f concava): $p_k = 1$ o $p_k = 0$ (cioé $\sum_{i=1}^{k-1} p_i = 1$)
  se $p_k = 1 \Rightarrow X$ concentrata in un'unico punto (cioé x_k)
- ($=$ in i.i.): $p_k = 0$ e, per i.i., X' concentrata in un'unico punto
  $\Rightarrow X$ concentrata in un'unico punto (tra $x_1 \ldots x_{k-1}$).

Proprietà dell'entropia

Lemma. Sia $X \sim P$ v.c. con alfabeto $\mathcal{X}$.

1. $H(X) \geq 0$; uguaglianza sse $\exists\, x \in \mathcal{X}$ t.c. $p(x) = 1$.
2. $H(X) \leq \lg |\mathcal{X}|$;
   l'uguaglianza vale sse X è uniformemente distribuita.

Dim. Punto 1.

- $0 \leq p(x) \leq 1 \quad \Rightarrow \log \frac{1}{p(x)} \geq 0 \Rightarrow \sum_x p(x) \log \frac{1}{p(x)} \geq 0$
- Esiste $x \in \mathcal{X}$ con $0 < p(x) < 1$ sse $H(x) > 0$

Dim. punto 2.

$$\begin{aligned} H(X) &= E\left(\lg \frac{1}{P(X)}\right) \\ &\leq \lg E\left(\frac{1}{P(X)}\right) \qquad \text{disug. Jensen su } \lg \text{ e } \begin{pmatrix} \dots & \frac{1}{p(x)} & \dots \\ \dots & p(x) & \dots \end{pmatrix} \\ &= \lg \sum_{x\in\mathcal{X}} p(x)\frac{1}{p(x)} = \lg \sum_{x\in\mathcal{X}} 1 = \lg|\mathcal{X}|. \end{aligned}$$

La disuguglianza di Jensen vale con il segno di “=” sse $\frac{1}{p(x)} = c$ costante, $\forall x \in \mathcal{X}$.

$\frac{1}{p(x)} = c \;\Rightarrow\; p(x) = \frac{1}{c} \;\Rightarrow\; 1 = \sum_{x\in\mathcal{X}} p(x) = \sum_{x\in\mathcal{X}} \frac{1}{c} = \frac{|\mathcal{X}|}{c}$.

Quindi $c = |\mathcal{X}|$ e $p(x) = \frac{1}{c} = \frac{1}{|\mathcal{X}|}$.

Condizionamento non aumenta entropia

$H(Y/X) \leq H(Y)$, uguaglianza sse X e Y indipendenti

$$
\begin{aligned}
H(Y/X) - H(Y) &= \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} p(x,y) \lg \frac{1}{p(y/x)} - \sum_{y \in \mathcal{Y}} p(y) \lg \frac{1}{p(y)} \\
&= \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} p(x,y) \lg \frac{1}{p(y/x)} - \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} p(x,y) \lg \frac{1}{p(y)} \\
&= \sum_{x,y} p(x,y) \lg \frac{p(y)}{p(y/x)} \leq \lg \sum_{x,y} p(x,y) \frac{p(y)}{p(y/x)} \\
&= \lg \sum_{x,y} \frac{p(y/x)p(x)p(y)}{p(y/x)} = \lg \sum_{x,y} p(x)p(y) \\
&= \lg \sum_{x} p(x) \sum_{y} p(y) = \lg \sum_{x} p(x) = \lg 1 = 0
\end{aligned}
$$

La disuguglianza di Jensen applicata a $\begin{pmatrix} \dots & \frac{p(y)}{p(y/x)} & \dots \\ \dots & p(x,y) & \dots \end{pmatrix}$

vale con il segno di “=” sse $\frac{p(y)}{p(y/x)} = c$ costante, $\forall x, y$.

$$\frac{p(y)}{p(y/x)} = c \qquad \Rightarrow \qquad p(y) = c\, p(y/x)$$

Sommando su y

$$1 = \sum_{y \in \mathcal{Y}} p(y) = \sum_{y \in \mathcal{Y}} c\, p(y/x) = c \sum_{y \in \mathcal{Y}} p(y/x) = c$$

Quindi $p(y) = p(y/x)$, $\forall x, y$, cioé X, Y sono indipendenti.

Corollario $H(X,Y) \leq H(X) + H(Y)$

Corollario
$H(X,Y/Z) = \sum_{x,y,z} p(xyz) \log \frac{1}{p(xy/z)} = H(X/Z) + H(Y/X,Z)$

Dim. (lasciata come esercizio)

Nota: In generale non è vero che $H(X/Y) = H(Y/X)$.

Estensione Regola della catena

Teorema Siano $X_1, \ldots, X_n$ n v.c. con d.d.p. $p(x_1, \ldots, x_n)$

$$H\left(X_1, ..., X_n\right) = \sum_{i=1}^{n} H\left(X_i / X_{i-1} ... X_1\right)$$

Dim.

Procediamo per induzione su n.

Per $n = 2$, applicando la regola della catena abbiamo

$$H\left(X_1, X_2\right) = H\left(X_1\right) + H\left(X_2 / X_1\right)$$

Iterando (assumiamo l'asserto vero per $n-1$)

$$\begin{aligned}
H\left(X_1,...,X_n\right) &= H\left(X_n/X_1,...,X_{n-1}\right)+H\left(X_1,...,X_{n-1}\right) \\
&= H(X_n/X_{n-1}\ldots X_1)+H(X_{n-1}/X_{n-2}\ldots X_1) \\
&\quad +\ldots+H(X_2/X_1)+H(X_1) \\
&= \sum_{i=1}^{n} H\left(X_i/X_{i-1},...,X_1\right)
\end{aligned}$$

$\square$

Teorema

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H\left(X_i\right),$$

l'uguaglianza vale sse $X_1, \ldots, X_n$ sono indipendenti.

Dim.

$$H\left(X_1, \ldots, X_n\right) = \sum_{i=1}^{n} H\left(X_i / X_1, ..., X_{i-1}\right) \leq \sum_{i=1}^{n} H\left(X_i\right)$$

L'ultima disuguaglianza vale con il segno di "=" sse $X_1, \ldots, X_n$ sono indipendenti. $\square$

Divergenza informazionale

Def. Date due d.d.p. $p(x)$ e $q(x)$,
la divergenza informazionale
(o *entropia relativa*, o *distanza Kullback Leibler*)
$D\left(p||q\right)$ di $p(x)$ e $q(x)$ è definita come

$$D\left(p||q\right) = \sum_{x \in \mathcal{X}} p(x) \lg \frac{p(x)}{q(x)}$$

Divergenza $D\left(p||q\right)$ misura la distanza tra le due d.p. p e q.
Non é simmetrica, infatti in generale $D\left(p||q\right) \neq D\left(q||p\right)$

$D(p||q) \geq 0$ **con l'uguaglianza sse** $p = q$

Proviamo $-D(p||q) \leq 0$ usando la disuguaglianza di Jensen:

$$\begin{aligned} -D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \lg \frac{q(x)}{p(x)} \leq \lg \sum_{x \in \mathcal{X}} p(x) \frac{q(x)}{p(x)} \\ &= \lg \sum_{x \in \mathcal{X}} q(x) = 0 \end{aligned}$$

dove dis. Jensen applicata a v.c. $\begin{pmatrix} \dots & \frac{q(x)}{p(x)} & \dots \\ \dots & p(x) & \dots \end{pmatrix}$:

Uguaglianza sse $\frac{q(x)}{p(x)} = c =$costante $\forall x$,
da cui $1 = \sum_x q(x) = c \sum_x p(x) = c$.
Quindi $c = 1$ e $q(x) = p(x)$, $\forall x$.

Date due v.c. X e Y con probabilità congiunta $p(x,y)$, la *mutua informazione* corrisponde all'entropia relativa tra $p(x,y)$ e $p(x)p(y)$:

$$
\begin{aligned}
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
&= \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} p(x,y) \lg \frac{p(x,y)}{p(x)\, p(y)} \\
&= D\left(p(x,y) || p(x)p(y)\right)
\end{aligned}
$$

Esempio

$Y \setminus X$	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Determiniamo le d.d.p. marginali.

Sommando le probabilità in ciascuna colonna:

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{pmatrix}$$

Sommando le probabilità in ciascuna riga:

$$Y = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$
\begin{aligned}
H\left(X/Y\right) &= \sum_{i=1}^{4} p\left(y=i\right) H\left(X/Y=i\right) \\
&= \frac{1}{4} H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{8}\right) + \frac{1}{4} H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{8}\right) \\
&\quad + \frac{1}{4} H\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right) + \frac{1}{4} H\left(1,0,0,0\right) \\
&= \frac{1}{4}\left[\frac{1}{2}+\frac{1}{4}\cdot 2 + 2\cdot\frac{1}{8}\cdot 3\right] 2 + \frac{1}{4}\lg 4 + \frac{1}{4}0 \\
&= \frac{1}{4}\left[\frac{1}{2}+\frac{1}{2}+\frac{3}{4}\right] 2 + \frac{1}{2} = \frac{1}{2}\left[\frac{7}{4}\right] + \frac{1}{2} = \frac{11}{8}
\end{aligned}
$$

$$H\left(Y/X\right) = \frac{13}{8}$$

$$H(X,Y) = H(Y) + H(X/Y) = 2 + \frac{11}{8} = \frac{27}{8}$$

$$\begin{aligned} I(X;Y) &= H(X) - H(X/Y) = \frac{7}{4} - \frac{11}{8} = \frac{3}{8} \\ &= H(Y) - H(Y/X) = 2 - \frac{13}{8} = \frac{3}{8} \end{aligned}$$

Esercizio

Esercizio Dimostrare che, data la funzione f, risulta

$$H\left(X\right) \geq H\left(f\left(X\right)\right)$$

Suggerimento: considerare $H\left(X, f\left(X\right)\right)$

Esempio

Condizioni meteo modellate da v.c. $X = \begin{pmatrix} x_p & x_n \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Yari esce con ombrello in accordo alla v.c.

$$Y = \begin{pmatrix} y_o & y_n \\ p(y_o) & p(y_n) \end{pmatrix} \text{ con } [p(y/x)] = \begin{array}{c|cc} X\backslash Y & y_o & y_n \\ \hline x_p & \frac{1}{3} & \frac{2}{3} \\ x_n & 1 & 0 \end{array}$$

Abbiamo $P(y_o) = p(x_p)p(y_o/x_p) + p(x_n)p(y_o/x_n) = 2/3$,

quindi $Y = \begin{pmatrix} y_o & y_n \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$

$$X = \begin{pmatrix} x_p & x_n \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad Y = \begin{pmatrix} y_o & y_n \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix},$$

$X\backslash Y$	y_o	y_n
x_p	$\frac{1}{3}$	$\frac{2}{3}$
x_n	1	0

Risulta

$H(X) = 1, \qquad H(Y) = h\left(\frac{1}{3}\right)$

$H(Y/X) = \sum_x p(x) H(Y/X = x) = \frac{1}{2} h\left(\frac{1}{3}\right) + \frac{1}{2} h(1) = \frac{1}{2} h\left(\frac{1}{3}\right)$

$I(X;Y) = H(Y) - H(Y/X) = h\left(\frac{1}{3}\right) - \frac{1}{2} h(\frac{1}{3}) = \frac{1}{2} h(\frac{1}{3})$

Zelda per indovinare che tempo fa osserva Yari:

- Se Yari ha ombrello dice x_p
- Se Yari non ha ombrello, lancia una moneta: se testa dice x_n, altr. dice x_p

Previsione Zelda dipende solo dal comportamento di Yari!

Sia Z la v.c. che rappresenta la previsione di Zelda.

Per ogni x, y, z

$$p(z/xy) = p(z/y)$$

Previsione Z(elda) fornisce meno informazione su X che l'osservazione diretta comportamento di Y(ari)!, Formalmente

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

Catene di Markov

X, Y, Z formano una catena di Markov ($X \rightarrow Y \rightarrow Z$)

sse

Z é condizionalmente indipendente da X noto Y, formalmente

$$p(z/xy) = p(z/y), \qquad \forall x, y, z$$

Teorema del Data Processing

Esercizio: $X \to Y \to Z$ sse $Z \to Y \to X$

Teorema(Data Processing) Se $X \to Y \to Z$ allora

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

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

Dim. Proviamo a).
Disuguaglianza b) segue applicando a) a $Z \to Y \to X$.

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

Poiché $X \to Y \to Z$

$$H(X/YZ) = E[-\lg p(x/yz)] = E[-\lg p(x/y)] = H(X/Y)$$

Da cui

$$I(X;YZ) = H(X) - H(X/YZ) = H(X) - H(X/Y) = I(X;Y)$$

Inoltre

$$I(X;YZ) = H(X) - H(X/YZ) \geq H(X) - H(X/Z) = I(X;Z)$$

Stima di v.c.

$$X = \begin{pmatrix} x_p & x_n \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \begin{array}{c|cc} Y/X & y_o & y_n \\ \hline x_p & \frac{1}{3} & \frac{2}{3} \\ x_n & 1 & 0 \end{array}, \quad Y = \begin{pmatrix} y_o & y_n \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Zelda per indovinare che tempo fa osserva Yari:

- Se Yari ha ombrello dice x_p
- Se Yari non ha ombrello dice x_n

v.c. $Z =$stima Zelda fa di X conoscendo Y.

$$Z = g(Y) = \begin{cases} x_p & \text{se } Y = y_o \\ x_n & \text{se } Y = y_n \end{cases}$$

Quale é la probabilità che Zelda fa una stima esatta di X?

Disuguaglianza di Fano

Siano X, Y v.c.. Vogliamo stimare X dall'osservazione di Y.
Osserviamo Y, calcoliamo $g(Y) = \hat{X}$ (stima di X).
Probabilitá $X \neq \hat{X}$? Sia

$$P_e = Pr\{\hat{X} \neq X\}$$

Teorema (Disuguaglianza di Fano)

$$h(P_e) + P_e \lg(|\mathcal{X}| - 1) \geq H(X|Y)$$

Nota: Teorema implica $P_e \geq \frac{H(X|Y)-h(P_e)}{\lg(|\mathcal{X}|-1)} \geq \frac{H(X|Y)-1}{\lg(|\mathcal{X}|)}$
Nota: $P_e = 0 \ \Rightarrow H(X|Y) = 0$
Nota: $X \to Y \to \hat{X}$.

Dim. . Definiamo la v.c. errore E t.c.

$$E = \begin{cases} 1 & \text{se } X \neq \hat{X} \\ 0 & \text{se } X = \hat{X} \end{cases}, \qquad \text{cioé} \qquad E = \begin{pmatrix} 1 & 0 \\ P_e & 1 - P_e \end{pmatrix}$$

$$\begin{aligned} H(X|Y) &= H(X|Y) + H(E|X,Y) = H(E,X|Y) \\ &= H(E|Y) + H(X|E,Y) \\ &\leq h(P_e) + H(X|E,Y) \end{aligned}$$

con

$$\begin{aligned} H(X|E,Y) &= Pr\{E=0\}H(X|E=0,Y) + Pr\{E=1\}H(X|E=1,Y) \\ &\leq (1-P_e)0 + P_e \lg(|\mathcal{X}|-1) = P_e \lg(|\mathcal{X}|-1) \end{aligned}$$

Riepilogo Misure di Informazione

Entropia: $H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}$

- Limiti: $0 \leq H(X) \leq \log |\mathcal{X}|$,
- Il condizionamento non aumenta l'entropia: $H(X/Y) \leq H(X)$
- Regola della catena:
  $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i/X_{i-1} \ldots x_1) \leq \sum_{i=1}^{n} H(X_i)$,
  $H(X_1, \ldots, X_n/Y_1, \ldots, Y_n) \leq \sum_{i=1}^{n} H(X_i/Y_i)$

Divergenza informazionale: $D\left(p||q\right) = \sum_{x \in \mathcal{X}} p(x) \lg \frac{p(x)}{q(x)} \geq 0$

Riepilogo Misure di Informazione

Mutua Informazione:

$$\begin{aligned} I(X;Y) &= H(X) - H(X/Y) = H(X) + H(Y) - H(X,Y) \\ &= D(p(x,y)||p(x)p(y)) \end{aligned}$$

- Simmetrica e positiva: $I(X;Y) = I(Y;X) \geq 0$,
- X,Y indipendenti $\Rightarrow I(X;Y) = H(X) - H(X/Y) = 0$
- $X \to Y \to Z \Rightarrow I(X;Z) \leq I(X;Y)$

Fano:

- $X \to Y \to \hat{X}$, $P_e = Pr\{\hat{X} \neq X\} \Rightarrow$
  $h(P_e) + P_e \lg(|\mathcal{X}| - 1) \geq H(X|Y)$