(F1X) Teoria dell'Informazione e della Trasmissione

Entropia

Docente: *Nicolò Cesa-Bianchi* versione 8 marzo 2016

Fissiamo un modello di sorgente $\langle \mathcal{X}, p \rangle$ con $\mathcal{X} = \{x_1, \ldots, x_m\}$ e $p = (p_1, \ldots, p_m)$. Introduciamo una funzione iniettiva $X : \mathcal{X} \to \{a_1, \ldots, a_m\}$ con $a_1, \ldots, a_m \in \mathbb{R}$ tali che $\mathbb{P}(X = a_i) = p_i$. Quindi X è una variabile casuale e la sua **entropia** è definita come

$$H(X) = \sum_{i=1}^{m} p_i \log_2 \frac{1}{p_i} \ .$$

Si noti che l'entropia è una proprietà della distribuzione $p_1, \ldots, p_m$ soltanto. In particolare, non dipende dal dominio $\{a_1, \ldots, a_m\}$ di X. Inoltre, dato che $0 \log 0 = 0$, l'entropia non dipende da quegli $x_i \in \mathcal{X}$ tali che $p(x_i) = 0$. L'unità di misura dell'entropia è il **bit** (Binary Information uniT).

Cambiare la base del logaritmo nel calcolo dell'entropia corrisponde a scalare la stessa per una costante positiva. Infatti, ricordiamo che $\log_b p = (\ln p)/(\ln b)$ per ogni $b > 1$ e $p > 0$. Questo implica $\log_b p = \frac{\ln p}{\ln b} \frac{\ln a}{\ln a} = \frac{\ln a}{\ln b} \frac{\ln p}{\ln a} = (\log_b a)(\log_a p)$. Quindi, introducendo la notazione $H_b(X)$ per l'entropia calcolata con $\log_b$,

$$H_b(X) = \sum_{i=1}^{m} p_i \log_b \frac{1}{p_i} = (\log_b a) \sum_{i=1}^{m} p_i \log_a \frac{1}{p_i} = (\log_b a) H_a(X) \ .$$

Un'altra proprietà dell'entropia è la non negatività, $H(X) \geq 0$. Infatti, l'entropia è una combinazione lineare con coefficienti positivi p_i di numeri non negativi $\log \frac{1}{p_i}$. In particolare, $H(X) = 0$ se e solo se esiste $a \in \mathbb{R}$ tale che $\mathbb{P}(X = a) = 1$. Infatti, $H(X) = 1 \ln 1 = 0$.

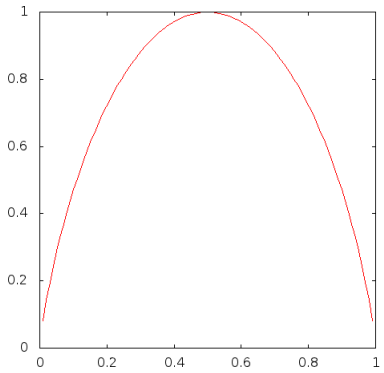

Figura 1: Grafico dell'entropia binaria $h(p)$ nel suo intervallo di definizione $[0, 1]$.

Un'entropia importante è quella binaria, ovvero l'entropia di una variabile casuale Bernoulliana $X \in \{0, 1\}$ dove $\mathbb{P}(X = 1) = p$ e $\mathbb{P}(X = 0) = 1 - p$ per $p \in [0, 1]$. Dalla definizione di entropia

abbiamo $H(x) = h(p)$ dove

$$h(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \ .$$

Si veda la Figura 1 per il grafico. Un'altra proprietà della funzione entropia è stabilita dal seguente semplice teorema. Per la dimostrazione utilizziamo le due disuguaglianze elementari $1 - \frac{1}{x} \leq \ln x \leq x - 1$ per ogni $x > 0$ facilmente dimostrabili via espansione in serie di Taylor. Si veda la Figura 2 per un'illustrazione grafica.

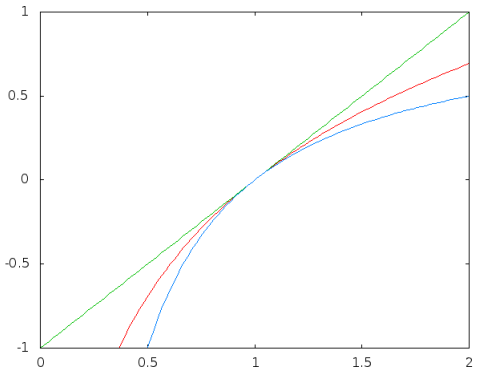

Figura 2: Illustrazione delle disuguaglianze $1 - \frac{1}{x} \leq \ln x \leq x - 1$ valida per ogni $x > 0$ (la linea rossa in mezzo è $\ln x$).

Fatto 1 *Sia X una variabile casuale che assume m valori distinti $a_1, \ldots, a_m$. Allora $H_D(X) \leq \log_D m$ per ogni $D > 1$. Inoltre, $H(X) = \log_D m$ se e solo se X ha distribuzione uniforme su $a_1, \ldots, a_m$.*

Dimostrazione. Utilizzando $\ln x \leq x - 1$ possiamo scrivere

$$\begin{aligned} H_d(X) - \log_D m &= \sum_{i=1}^{m} p_i \log_D \frac{1}{p_i} - \log_D m = \sum_{i=1}^{m} p_i \log_D \frac{1}{m\, p_i} = \sum_{i=1}^{m} p_i \left(\ln \frac{1}{m\, p_i} \right) \frac{1}{\ln D} \\ &\leq \sum_{i=1}^{m} p_i \left(\frac{1}{m\, p_i} - 1 \right) \frac{1}{\ln D} = \left(1 - \sum_{i=1}^{m} p_i \right) \frac{1}{\ln D} = 0 \ . \end{aligned}$$

Quindi abbiamo dimostrato che $H_D(X) \leq \log_D m$. Ora, se $\mathbb{P}(X = a_i) = \frac{1}{m}$ per $i = 1, \ldots, m$ abbiamo che

$$H_D(X) = \sum_{i=1}^{m} \frac{1}{m} \log_D m = \log_D m \ .$$

□

Dimostriamo ora un importante risultato che lega l'entropia alla lunghezza media delle parole di un codice istantaneo. Prima però dobbiamo introdurre una quantità che ci servirà nella dimostrazione

e che ha un significato importante di per sè. Consideriamo due variabili casuali X e Y definite su uno stesso dominio $\mathcal{S}$. Sia p_X la distribuzione di X e p_Y quella di Y. L'**entropia relativa** (in base $D > 1$) fra X e Y è definita come

$$D(X\|Y) = \sum_{s\in\mathcal{S}} p_X(s) \log_D \frac{p_X(s)}{p_Y(s)}$$

dove $0 \log_D \frac{0}{0} = 0 \log_D 0 - 0 \log_D 0 = 0$.

L'entropia relativa misura la diversità fra due distribuzioni di probabilità. Non è una distanza, infatti è asimmetrica. Cioè, in generale, $D(X\|Y) \neq D(Y\|X)$. Inoltre, può assumere valori illimitati. Infatti, se esiste $s \in \mathcal{S}$ tale che $p_X(s) > 0$ e $p_Y(s) = 0$ allora $D(X\|Y) = \infty$, come di deduce dalla definizione di entropia relativa. D'altra parte, $D(X\|X) = 0$ per ogni variabile casuale X, come ancora si può dedurre dalla definizione.

Il seguente risultato, chiamato *information inequality*, mostra che —come l'entropia— l'entropia relativa è non negativa.

Teorema 2 *Per ogni coppia di variabili casuali X, Y definite su un comune dominio $\mathcal{S}$ vale la disuguaglianza $D(X\|Y) \geq 0$.*

Dimostrazione. Utilizzando la disuguaglianza $\ln x \geq 1 - \frac{1}{x}$ possiamo scrivere

$$\begin{aligned} D(X\|Y) &= \sum_{s\in\mathcal{S}} p_X(s) \log_D \frac{p_X(s)}{p_Y(s)} = \sum_{s\in\mathcal{S}} p_X(s) \left(\ln \frac{p_X(s)}{p_Y(s)}\right) \frac{1}{\ln D} \\ &\geq \frac{1}{\ln D} \sum_{s\in\mathcal{S}} p_X(s) \left(1 - \frac{p_Y(s)}{p_X(s)}\right) = \frac{1}{\ln D}(1-1) = 0 \ . \end{aligned}$$

□

Siamo pronti a dimostrare il risultato principale di questa sezione.

Teorema 3 *Se $c : \mathcal{X} \to \mathcal{D}^+$ è un codice istantaneo D-ario per $\langle \mathcal{X}, p\rangle$ allora $\mathbb{E}\big[\ell_c\big] \geq H_D(X)$.*

Dimostrazione. Sia $Z : \mathcal{X} \to \mathbb{R}$ una variabile casuale con distribuzione $q(x) = D^{-\ell_c(x)} \Big/ \Big(\sum_{x'} D^{-\ell_c(x')}\Big)$.

Possiamo quindi scrivere

$$\begin{aligned}
\mathbb{E}\big[\ell_c\big] - H_D(X) &= \sum_{x\in\mathcal{X}} p(x)\ell_c(x) - \sum_{x\in\mathcal{X}} p(x)\log_D \frac{1}{p(x)} \\
&= \sum_{x\in\mathcal{X}} p(x)\left(\log_D D^{\ell_c(x)} + \log_D p(x)\right) \\
&= \sum_{x\in\mathcal{X}} p(x)\left(\log_D \frac{p(x)}{D^{-\ell_c(x)}}\frac{\sum_{x'} D^{-\ell_c(x')}}{\sum_{x'} D^{-\ell_c(x')}}\right) \\
&= \sum_{x\in\mathcal{X}} p(x)\left(\log_D p(x)\frac{\sum_{x'} D^{-\ell_c(x')}}{D^{-\ell_c(x)}} - \log_D \sum_{x'} D^{-\ell_c(x')}\right) \\
&= \underbrace{D(X\|Z)}_{\geq 0} - \log_D \underbrace{\sum_{x'} D^{-\ell_c(x')}}_{\leq 1} \geq 0
\end{aligned}$$

dove abbiamo usato il Fatto 2 per $D(X\|Z) \geq 0$ e la disuguaglianza di Kraft nell'ultimo passaggio, che sfrutta l'ipotesi che il codice c è istantaneo. □

Questo risultato dà un signiicato operativo all'entropia, mostrando che l'entropia di una variabile casuale X è un limite inferiore al numero di simboli di codice istantaneo che dobbiamo usare in media per descrivere una realizzazione di X.