Soluzioni - foglio 7

8 maggio 2023

E7.1 Per quale motivo non possono esistere insiemi X e Y tali che $H(X) = 3$, $H(Y) = 4$ e $H(X, Y) = 8$? Che cosa puoi dire di X e Y se, invece, $H(X, Y) = 7$?
Soluzione:
Perché per ogni X e Y si ha che $H(X, Y) \leq H(X) + H(Y)$; il segno di uguaglianza si ha solo nel caso in cui X e Y sono indipendenti.

E7.2 Siano X e Y due variabili casuali indipendenti con $H(X) = 11$ e $H(Y) = 2$. Determina l'entropia congiunta $H(X, Y)$ e la mutua informazione $I(X, Y)$.
Soluzione:
Dato che si tratta di variabili indipendenti, si ha che $I(X, Y) = 0$ e $H(X, Y) = H(X) + H(Y)$. In generale, valgono le seguenti quattro relazioni (vedi anche la Figura 3.2 delle note)

$$\begin{cases} H(X,Y) = H(X) + H(Y|X) \\ H(X,Y) = H(Y) + H(X|Y) \\ H(Y) = I(X,Y) + H(Y|X) \\ H(X) = I(X,Y) + H(X|Y) \end{cases}$$

E7.3 Se $X = \{a, b, c, d, e, f\}$ con $p(a) = p(b) = p(c) = p(d) = 1/8$ e $p(e) = p(f) = 1/4$, calcola $H(X)$ e la lunghezza media della codifica C con $C(a) = 100$, $C(b) = 101$, $C(c) = 110$, $C(d) = 111$, $C(e) = 00$ e $C(f) = 01$. Discuti la decifrabilità e l'istantaneità di C. Perchè non può esistere una codifica più efficiente di C?
Soluzione:
Entropia:

$$\begin{aligned} H(X) &= \sum_{x \in \mathcal{X}} p(x) \log_2 p(x)^{-1} \\ &= 4\frac{1}{8} \log_2 8 + 2\frac{1}{4} \log_2 4 = \frac{5}{2}. \end{aligned}$$

Lunghezza media codifica:

$$L(C, X) = \sum_{x \in \mathcal{X}} p(x) L_C(x) = \frac{5}{2}.$$

La codifica è univocamente decifrabile in quanto la codifica estesa è iniettiva. La codifica è istantanea perché nessuna parola della codifica è prefisso di un'altra parola della codifica.

E7.4 Calcola la codifica di Huffman per i simboli dell'esercizio precedente.
Soluzione:
Una possibile codifica è: C con $C = [C(a) = 111, C(b) = 011, C(c) = 110, C(d) = 010, C(e) = 00, C(f) = 10]$.

E7.5 Sia dato l'insieme di simboli $X = a, b, c$ e la codifica $C(a) = 0$, $C(b) = 01$, $C(c) = 001$. Stabilire (giustificando opportunamente) se la codifica: 1) è univocamente decifrabile; 2) è istantanea; 3) soddisfa la disuguaglianza di Kraft-McMillian. Se $\mathcal{X}$ sono i simboli di una sorgente X, cosa si può dire rispetto all'entropia $H(X)$ e alla lunghezza media delle parole del codice C?
Soluzione:

1 La codifica non è univocamente decifrabile, la sequenza ab è indstinguibile dalla sequenza c.

2 La codifica non è istantanea, a è prefisso di b e c.

3 Sì, soddisfa la disuguaglianza di Kraft-McMillian $\sum_x 2^{-L(x)} = 0.875$.

4 Il primo teorema di Shannon dimostra la seguente diseguaglianza: $\bar{L}(X) \geq H(X)$. Dove $\bar{L}(x)$ e' la lunghezza media delle parole del codice C.

E7.6 Data una codifica aritmetica in cui la probabilità di ogni simbolo è fissata a priori spiega perché ti aspetti codifiche più brevi per sequenze corrispondenti a intervalli di ampiezza più grandi.
Soluzione:
Perchè nella codifica aritmetica l'ampiezza dell'intervallo è data dal prodotto delle probabilità di tutti i simboli che compongono il messaggio.