Esercizi - Informazione 2

E8.1 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 $I(X,Y)$ (vedere la fine della sezione 3.15 dell'ultima versione delle note del corso su Aulaweb).

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)

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

E8.2 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**. Perché non può esistere una codifica più efficiente di **C**?

Soluzione:
Entropia:

$$\begin{aligned} H(X) &= \sum_{x\in\mathcal{X}} p_x \log_2 \frac{1}{p_X(x)} \\ &= 4 \cdot \frac{1}{8} \log_2 8 + 2 \cdot \frac{1}{4} \log_2 4 \\ &= \frac{5}{2} \end{aligned}$$

Lunghezza media codifica:

$$\begin{aligned} L(C,X) &= \sum_{x\in\mathcal{X}} p_x L_C(x) \\ &= 4 \cdot 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} \\ &= \frac{11}{4} = 2.75 \end{aligned}$$

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.

E8.3 Calcola la codifica di Huffman per i simboli dell'esercizio precedente.

Soluzione:
Una possibile codifica è: **C** con $\mathbf{C} = \left[\mathbf{C}(a) = 111, \mathbf{C}(b) = 011, \mathbf{C}(c) = 110, \mathbf{C}(d) = 010, \mathbf{C}(e) = 00, \mathbf{C}(f) = 10\right]$

E8.4 Sia dato l'insieme di simboli $\mathcal{X} = \{a, b, c\}$ 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* $\bar{L}(\mathbf{C})$ *è la lunghezza media delle parole del codice* $\mathbf{C}$.

E8.5 Sia dato $X = \{x_1, x_2, x_3, x_4\}$ e l'insieme di interi $\{L_1, L_2, L_3, L_4\}$ con $L_1 = 1$, $L_2 = 2$, $L_3 = 2$ e $L_4 = 3$. Per quale motivo non può esistere una codifica istantanea C che abbia gli interi L_i come lunghezze delle rappresentazioni $C(x_i)$?

Soluzione:
Perché $2^{-1} + 2^{-2} + 2^{-2} + 2^{-3} = 1.125 > 1$.

E8.6 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:
Perché nella codifica aritmetica l'ampiezza dell'intervallo è data dal prodotto delle probabilità di tutti i simboli che compongono il messaggio.