UNIVERSITÀ DEGLI STUDI "ROMA TRE"
CORSO DI STUDI IN MATEMATICA
IN420 - TEORIA DELL'INFORMAZIONE – A.A. 2015-2016

M. PEDICINI

ESAME DEL 20/01/2016 – TEMPO 3H00

COGNOME ________________________ NOME ________________________ MATRICOLA __________________

Esercizio 1. *Si consideri una variabile aleatoria con k valori, e siano p_i, per $i = 1, \ldots, k$ le rispettive probabilità ordinate in modo che $p_i < p_{i+1}$.*

Si costruisca una codifica in accordo con la seguente procedura: si divide l'insieme in due, ad un certo indice m, in modo che le somme dei due gruppi che si formano siano il più vicino possibile tra di loro, ovvero in modo da minimizzare

$$\sum_{i=1}^{m} p_i - \sum_{i=m+1}^{k} p_i$$

e si assegni al primo gruppo il bit 0 e al secondo gruppo il bit 1 e si iteri la procedura sui sottogruppi così formatisi.

Questo metodo viene spesso indicato come codice di Fano.

Per ognuna delle seguenti distribuzioni trovare il codice di Fano e dire se coincide con il codice Huffman e se sia oppure no ottimale.

$$p_a = \{4/10, 3/10, 2/10, 1/10\}$$
$$p_b = \{6/21, 5/21, 4/21, 3/21, 2/21, 1/21\}$$
$$p_c = \{15/43, 7/43, 7/43, 7/43, 7/43\}$$

Esercizio 2. *Siano X e Y due variabile aleatorie discrete sullo stesso alfabeto $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$. che differesicono per le due distribuzioni di probabilità su $\mathcal{A}$ siano p_X e p_Y.*

Consideriamo due schemi di codifica $C_1(x)$ e $C_2(x)$ per gli elementi di $\mathcal{A}$:

x	$p_X(x)$	$p_Y(x)$	$C_1(x)$	$C_2(x)$
a_1	1/2	1/2	0	0
a_2	1/4	1/8	10	100
a_3	1/8	1/8	110	101
a_4	1/16	1/8	1110	110
a_5	1/16	1/8	1111	111

(1) *Calcolare $H(X)$, $H(Y)$, e le entropie relative (distanze Kullback-Leibler) $D(p_X||p_Y)$ e $D(p_Y||p_X)$.*
(2) *Mostrare che la lunghezza media per le parole codice di C_1 utilizzata con X coincide con $H(X)$ e pertanto la codifica C_1 è ottimale per X. Mostrare che C_2 è ottimale per Y.*
(3) *Si supponga di utilizzare la codifica C_2 per X. Che lunghezza media otteniamo per le parole codice in questo caso ? Di quanto è maggiore di $H(X)$? Si evidenzi in relazione tra la risposta e il valore di $D(p_X||p_Y)$.*
(4) *Se utilizziamo C_1 su Y di quanto la lunghezza media delle parole codice differisce da $H(Y)$? Che relazione c'è con $D(p_Y||p_X)$?*

Esercizio 3. *Codificare con l'algoritmo Lempel-Ziv il seguente vettore binario:*

110 011 100 100 011 100 001 100 000 110 011 000.

Esercizio 4. *Si consideri un canale di comunicazione asimmetrico il cui alfabeto di ingresso sia binario* $\mathcal{A}_X = \{0,1\}$ *con distribuzione uniforme* $p_X(x=0) = \dfrac{1}{2}$ *e in cui l'alfabeto di uscita* $\mathcal{A}_Y = \{0,1,2\}$ *abbia tre elementi, la distribuzione sull'alfabeto di uscita è determinata dal canale descritto nel modo seguente: l'input* 0 *è ricevuto correttamente con probabilità* $1-\alpha$*, e l'input* 1 *viene ricevuto correttamente con probabilità* $1-\beta$*, inoltre l'input* 0 *viene scambiato con uno degli altri due simboli con probabilità uniforme* $\alpha/2$*, mentre il simbolo* 1 *viene ricevuto come* 0 *con probabilità* $\beta/3$ *e come* 2 *con probabilità* $2\beta/3$*; risultando così in una matrice di transizione per il canale*

$$p(y_k|x_j) = \begin{pmatrix} 1-\alpha & \beta/3 \\ \alpha/2 & 1-\beta \\ \alpha/2 & 2\beta/3 \end{pmatrix}.$$

(1) *Determinare la distribuzione sull'alfabeto di uscita* p_Y.
(2) *Determinare i valori di* α *e* β *che massimizzano la mutua informazione ottenuta sul canale, e calcolare la capacità del canale.*
(3) *Calcolare i valori di* α *e* β *che minimizzano la mutua informazione sul canale.*