INFORMACIÓ I SEGURETAT
26 de març de 2019

Nom i cognoms (en MAJÚSCULES): ______________________________ NIU: __________ Grup: ___

- Cal que **justifiqueu convenientment** totes les respostes.
- $\log 3 = 1.58$, $\log 5 = 2.32$, $\log 7 = 2.80$, $\log 23 = 4.52$.

1. (1.5 punt, 0.5+0.5+0.5) Una persona escull a l'atzar de manera equiprobable una paraula del conjunt $S = \{\text{GOL, PAS, GEL, PES}\}$. **Nota:** Justifiqueu les respostes en termes de teoria de la informació.

   (a) Quina incertesa mitjana tenim sobre quina és la segona lletra de la paraula escollida?

   (b) Quina informació mitjana ens dóna la segon lletra si coneixem la primera?

   (c) Quantes preguntes de mitjana amb resposta SI/NO hem de fer per encertar la paraula escollida?

   Solució:

   (a) Considerem els conjunts S_i que contenen les lletres de les paraules en la posició i, per $i \in \{1, 2, 3\}$. Així, $S_2 = \{O, A, E\}$ amb probabilitats $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\}$ Per tant, la incertesa mitjana que tenim sobre quina és la segona lletra de la paraula escollida és $H(S_2) = H(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}) = 2\frac{1}{4}\log(4) + \frac{1}{2}\log(2) = 1.5$ bits.

   (b) Si coneixem la primera, la probabilitat de que la segona sigui O, A o E bé donada per la següent taula:

$p((S_2)_j \mid (S_1)_i)$	O	A	E
G	$\frac{1}{2}$	0	$\frac{1}{2}$
P	0	$\frac{1}{2}$	$\frac{1}{2}$

   Tenim $H(S_2 \mid S_1 = G) = H(\frac{1}{2}, \frac{1}{2}) = 1$ i $H(S_2 \mid S_1 = P) = H(\frac{1}{2}, \frac{1}{2}) = 1$. Per tant, $H(S_2|S_1) = 1$ bit.

   (c) La font consta de 4 esdeveniments que són equiprobables i, per tant, té una entropia de $\log 4 = 2$ bits. Cada resposta ens aporta una informació màxima de 1 bit ($\log 2$) si realitzem una pregunta on les possibles respostes SI/NO són equiprobables. Per tant, el nombre de preguntes que hem de fer és 2.

2. (2.5 punts, 0.5+1+0.5+0.5)

   (a) Considereu la font $S_1 = \{a_1, a_2, a_3, a_4\}$, amb $p(a_4) = \frac{1}{3}$.

      i. Quina és l'entropia màxima que pot tenir aquesta font?

      ii. Pot existir un codi binari òptim amb longitud mitjana igual a aquesta entropia? En cas afirmatiu, doneu el codi. En cas negatiu, digueu quina és la menor longitud mitjana possible i doneu un codi amb aquesta longitud mitjana.

   (b) Considereu ara la font $S_2 = \{a_1, \ldots, a_n\}$, amb $H(S) = 3.16$.

      i. Quin és el valor mínim de n?

      ii. Sigui C un codi ternari instantani amb $\bar{L} = 3.16$ per codificar els elements de S_2. Pot ser òptim?

   Solució:

   (a) i. L'entropia màxima es dóna quan $p(a_1) = p(a_2) = p(a_3) = \frac{2}{9}$ i, per tant, és $H(\frac{2}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{3}) = \frac{6}{9}\log(3) - \frac{6}{9}\log(2) + \log(3) = \frac{5}{3}1.58 - \frac{2}{3} = \frac{7.9}{3} - \frac{2}{3} = \frac{5.9}{3} = 1.96$.

ii. No, ja que les probabilitats no són potències de dos (pel Primer Teorema de Shannon). Si apliquem l'algorisme de Huffman, obtenim el següent codi binari òptim $C = \{11, 10, 01, 00\}$ i la longitud mitjana és 2.

(b) i. Sabem que $H(S) \leq \log(n)$. Per tant, $3.16 \leq \log(n)$, que implica $2^{3.16} \leq n$ i així $n \geq 9$.

ii. Sabem que existeix un codi ternari instantani tal que $\bar{L} \leq \frac{H(S)}{log(3)} + 1 = \frac{3.16}{1.58} + 1 = 3$. Per tant, si tenim un codi ternari instantani amb $\bar{L} = 3.16$, no pot ser òptim.

3. (2.5 punts, 0.5+0.5+1.5) Volem enviar a través d'un canal binari, apostes a quinieles; és a dir, cadenes de longitud 15 amb símbols del conjunt $S = \{1, 2, X\}$.

(a) Si volem codificar els símbols de S amb un codi binari de longitud fixa, quina és la longitud del codi?

(b) I si volem codificar triples de símbols (símbols de 3 en 3)?

(c) Descodifiqueu la seqüència (0,'X')(1,'1')(2,'2')(0,'2')(3,'1')(3,'X') que s'ha comprimit fent servir el mètode de compressió LZ78. Doneu el percentatge de compressió si cada índex ocupa 2 bits i la mida de bits per representar cada símbol és l'obtinguda a l'apartat (a).

Solució:

(a) La longitud ha complir $2^L \leq 3$. Per tant, $L = 2$.

(b) Si els volem codificar de 3 en 3, aleshores $2^L \leq 3^3 = 27$. Així, $L = 5$.

(c) .

Entrada	Pos.	Dicc.
	0	Null
(0,'X')	1	X
(1,'1')	2	X1
(2,'2')	3	X12
(0,'2')	4	2
(3,'1')	5	X121
(3,'X')	6	X12X

La cadena original és XX1X122X121X12X. La mida de la cadena original és $|M| = 15 \cdot 2 = 30$ bits, i la de la cadena comprimida $|C| = 6(2+2) = 24$ bits. La taxa de compressió és $R = \frac{24}{30} = \frac{4}{5}$ i el percentatge de compressió $(1 - \frac{4}{5}) \cdot 100 = \frac{100}{5} = 20\%$.

4. (3.5 punts, 0.5+0.5+1.5+0.5+0.5) Sigui $\{A_1, A_2, A_3, A_4\}$ el conjunt d'entrades i $\{B_1, B_2, B_3, B_4\}$ el conjunt de sortides d'un canal, amb matriu de probabilitats condicionades:

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

(a) De quin tipus de canal es tracta? Doneu la capacitat del canal.

(b) Quina distribució inicial fa que $H(B)$ sigui màxima?

(c) Considereu la distribució inicial $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{3}{8}\}$.

i. Quina és la informació mútua entre A i B?
ii. Quina és la regla a mínima probabilitat d'error (MPE)?
iii. Quina és la probabilitat mitjana d'error fent servir la regla de l'apartat anterior?

Solució:

(a) És un canal totalment simètric. La capacitat del canal és $C = \log(4) - H(\frac{1}{2}, \frac{1}{2}, 0, 0) = 1$ bit.

(b) Tenim que $H(B)$ és màxima si la distribució final és equiprobable. En un canal totalment simètric, això es dóna quan la distribució inicial és equiprobable. Per tant la distribució inicial és $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$.

(c) i. La informació mútua entre A i B és $I(A, B) = H(B) - H(B \mid A)$. Per ser un canal totalment simètric, $H(B \mid A) = H(\frac{1}{2}, \frac{1}{2}, 0, 0) = 1$.
Si la distribució de probabilitats inicial és $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{3}{8}\}$, aleshores la matriu de probabilitats conjuntes és la següent:

$$\begin{pmatrix} \frac{1}{8} & 0 & 0 & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & 0 & 0 \\ 0 & \frac{1}{16} & \frac{1}{16} & 0 \\ 0 & 0 & \frac{3}{16} & \frac{3}{16} \end{pmatrix}.$$

i les probabilitats $P(B_i)$ són, respectivament, $\frac{1}{4}, \frac{3}{16}, \frac{1}{4}, \frac{5}{16}$, per tant

$$H(B) = \frac{1}{4}\log(4) + \frac{3}{16}(\log(16) - \log(3)) + \frac{1}{4}\log(4) + \frac{5}{16}(\log(16) - \log(5)) = 1.977$$

Finalment, $I(A, B) = 1.977 - 1 = 0.977$.

ii. Fixant-nos en els valors màxims a cada columna de la matriu anterior obtenim la següent funció de descodificació a mínima probabilitat d'error:

$$\begin{array}{lcl} B_1 & \longrightarrow & A_1(\text{ o } A_2) \\ B_2 & \longrightarrow & A_2 \\ B_3 & \longrightarrow & A_4 \\ B_4 & \longrightarrow & A_4 \end{array}$$

iii. La probabilitat mitjana d'error en la descodificació és

$$\overline{p}_e = 1 - \frac{1}{8} - \frac{1}{8} - \frac{3}{16} - \frac{3}{16} = 1 - \frac{10}{16} = \frac{3}{8} = 0.375.$$