Τμήμα
Μηχανικών
Πληροφορικής τ.ε.
Τεχνολογικό Εκπαιδευτικό Ίδρυμα
Δυτικής Ελλάδας

Εξέταση μαθήματος
«Θεωρία Πληροφορίας»
Αντίρριο 9/9/2016
Εξεταστής: Μιχάλης Παρασκευάς

Ονοματεπώνυμο: ……………………………………………………………… ΑΜ ……………… [ΕΥΑΝΑΓΝΩΣΤΑ !!!]

Θέμα 1ο (3 μονάδες)

Δίνονται οι τυχαίες μεταβλητές X και Y, με δύο δυνατά αποτελέσματα η κάθε μία. Οι δισδιάστατες πιθανότητες είναι:

$$p(x_1,\, y_1) = \frac{1}{4}, \quad p(x_1,\, y_2) = \frac{1}{8}, \quad p(x_2,\, y_1) = \frac{1}{4}, \quad p(x_2,\, y_2) = \frac{1}{2}$$

1. Ποια ποσότητα πληροφορίας λαμβάνουμε όταν μας γνωστοποιείται το αποτέλεσμα της τυχαίας μεταβλητής X; Ποιά όταν μας γνωστοποιείται το αποτέλεσμα της Y; **[1 μονάδα]**

   Αρχικά θα υπολογίσουμε τις πιθανότητες $p(x_1),\ p(x_2),\ p(x_3),\ p(x_4)$. Είναι:

$$\boldsymbol{p(x_1) = \sum_{j=1}^{2} p(x_1,\, y_j) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} \qquad p(x_2) = \sum_{j=1}^{2} p(x_2\, y_j) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}}$$

$$\boldsymbol{p(y_1) = \sum_{i=1}^{2} p(x_i,\, y_1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \qquad p(y_2) = \sum_{i=1}^{2} p(x_i,\, y_2) = \frac{1}{8} + \frac{1}{2} = \frac{5}{8}}$$

   Κατόπιν υπολογίζουμε:

$$\boldsymbol{H(X) = -\sum_{i=1}^{2} p(x_i)\ log_2\ p(x_i) = -\left[\frac{3}{8} log_2\left(\frac{3}{8}\right) + \frac{3}{4} log_2\left(\frac{3}{4}\right)\right] = 0,8419\ bits}$$

$$\boldsymbol{H(Y) = -\sum_{i=1}^{2} p(y_i)\ log_2\ p(y_i) = -\left[\frac{1}{2} log_2\left(\frac{1}{2}\right) + \frac{5}{8} log_2\left(\frac{5}{8}\right)\right] = 0,9238\ bits}$$

2. Ποια είναι η ποσότητα πληροφορίας που λαμβάνουμε όταν γνωρίζουμε το αποτέλεσμα της συνδυασμένης διαδικασίας (X,Y); **[1 μονάδα]**

$$\boldsymbol{H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,\, y_i)\ log\ p(x_i,\, y_i)}$$
$$\boldsymbol{= -\left[\frac{1}{4} log_2\left(\frac{1}{4}\right) + \frac{1}{8} log_2\left(\frac{1}{8}\right) + \frac{1}{4} log_2\left(\frac{1}{4}\right) + \frac{1}{2} log_2\left(\frac{1}{2}\right)\right] = 1,875\ bits}$$

3. Ποια ποσότητα πληροφορίας λαμβάνουμε όταν μας γνωστοποιείται το αποτέλεσμα της Y αν γνωρίζουμε το X; **[1 μονάδα]**

$$\boldsymbol{H(X,Y) = H(X) + H(Y/X) \Rightarrow H(Y/X) = H(X,Y) - H(X) = 1,875 - 0,8419 = 1,0331\ bits}$$

Θέμα 2ο (4 μονάδες)

Μια διακριτή πηγή πληροφορίας χωρίς μνήμη παράγει 8 σύμβολα Α, Β, Γ, Δ, Ε, Ζ, Η και Θ με τις πιθανότητες $\left\{\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{4}, \frac{1}{32}, \frac{1}{8}\ \kappa\alpha\iota \frac{1}{8}\right\}$ αντίστοιχα.

1) Να βρεθούν τα σύμβολα με το χαμηλότερο και το υψηλότερο πληροφοριακό περιεχόμενο, αντίστοιχα **[1 μονάδα]**

Το σύμβολο με την υψηλότερη πιθανότητα εμφάνισης έχει το χαμηλότερο πληροφορικό περιεχόμενο. Στην προκειμένη περίπτωση είναι τα σύμβολα Α και Ε με πιθανότητα εμφάνισης 1/4, για τα οποία ισχύει $\boldsymbol{H(A) = H(E) = -log(1/4) = 2\ bits.}$

Το σύμβολο με την χαμηλότερη πιθανότητα εμφάνισης έχει το υψηλότερο πληροφορικό περιεχόμενο. Στην προκειμένη περίπτωση είναι τα σύμβολα Δ και Ζ με πιθανότητα εμφάνισης 1/32, για τα οποία ισχύει $\boldsymbol{H(\Delta) = H(Z) = -log(1/32) = 5\ bits.}$

2) Να βρεθεί το μέσο πληροφοριακό περιεχόμενο των συμβόλων της πηγής **[1 μονάδα]**

Το μέσο πληροφορικό περιεχόμενο των συμβόλων της πηγής υπολογίζεται από τη σχέση:

$$\boldsymbol{H(S) = -\sum_{i=1}^{n} p_i\, log_2\, p_i = \sum_{i=1}^{n} p_i log_2\left(\frac{1}{p_i}\right) =}$$

$$\boldsymbol{= \frac{1}{4} log_2 4 + \frac{1}{8} log_2 8 + \frac{1}{16} log_2 16 + \frac{1}{32} log_2 32 + \frac{1}{4} log_2 4 + \frac{1}{32} log_2 32 + \frac{1}{8} log_2 8 + \frac{1}{8} log_2 8}$$

$$\boldsymbol{= \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{5}{32} + \frac{2}{4} + \frac{5}{32} + \frac{3}{8} + \frac{3}{8} = \frac{43}{16} = 2,6875\ bits}$$

3) Να βρεθεί ο πλεονασμός της πηγής **[1 μονάδα]**

$$\boldsymbol{red = 1 - \frac{H(S)}{max\, H(S)} = 1 - \frac{H(S)}{log_2 8} = 1 - \left(\frac{2,6875}{3}\right) = 1 - 0,8958 = 0,1042}$$

4) Να βρεθεί ο μέσος ρυθμός πληροφορίας της πηγής για ρυθμό συμβόλων 10.000 σύμβολα/sec **[1 μονάδα]**

$$\boldsymbol{R = rH(S) = 10.000\ x\ (2,6875) = 26.875\ bits/sec}$$

<u>Θέμα 3ο (4 μονάδες)</u>

Δίδεται μία διακριτή πηγή πληροφορίας χωρίς μνήμη που παράγει 7 διαφορετικά σύμβολα U1…U7 με πιθανότητες αντίστοιχα {0.20, 0.15, 0.10, 0.30, 0.06, 0.15, 0.04}. Ζητείται:

1) Να σχεδιασθεί δυαδικός κώδικας σύμφωνα με τον αλγόριθμο Huffman. Εξηγήστε τα βήματα που ακολουθήσατε. **[1,5 μονάδα]**

Διατάσσουμε τα σύμβολα με αύξουσα σειρά πιθανότητας και εφαρμόζουμε τον αλγόριθμο Huffman, δηλ. σε κάθε βήμα αθροίζουμε τις δύο μικρότερες πιθανότητες. Σε κάθε «διακλάδωση» αναθέτουμε τα σύμβολα 1 και 0 (δεν φαίνονται στο παρακάτω σχήμα).

U7	**U5**	**U3**	**U2**	**U6**	**U1**	**U4**
0,04	0,06	0,10	0,15	0,15	0,20	0,30
	0,10	0,10	0,15	0,15	0,20	0,30
		0,20	0,15	0,15	0,20	0,30
		0,20		**0,30**	0,20	0,30
				0,30	**0,40**	0,30
					0,40	**0,60**
						1,00

Επομένως ο κώδικας κατά Huffman είναι: U7 = 1111 U5 = 1110
U3 = 110 U2 = 011 U6 = 010 U1 = 10 U4 = 00

2) Να σχεδιασθεί δυαδικός κώδικας σύμφωνα με τον αλγόριθμο Fano. Εξηγήστε τα βήματα που ακολουθήσατε. **[1,5 μονάδα]**

Ταξινομούμε τα σύμβολα U1…U7 κατά φθίνουσα πιθανότητα και χωρίζουμε σε ισοπίθανες ή σχεδόν ισοπίθανες ομάδες. Στο τέλος αναθέτουμε τα ψηφία ανά σύμβολο.

<table>
<tr><td>U4</td><td>0.3</td><td rowspan="2">0</td><td>0</td><td></td><td></td><td>00</td></tr>
<tr><td>U1</td><td>0.2</td><td>1</td><td></td><td></td><td>01</td></tr>
<tr><td>U6</td><td>0.15</td><td rowspan="5">1</td><td rowspan="2">0</td><td>0</td><td></td><td>100</td></tr>
<tr><td>U2</td><td>0.15</td><td>1</td><td></td><td>101</td></tr>
<tr><td>U3</td><td>0.1</td><td rowspan="3">1</td><td>0</td><td></td><td>110</td></tr>
<tr><td>U5</td><td>0.06</td><td rowspan="2">1</td><td>0</td><td>1110</td></tr>
<tr><td>U7</td><td>0.04</td><td>1</td><td>1111</td></tr>
</table>

Επομένως ο κώδικας κατά Fano είναι: U7 = 1111 U5 = 1110
U3 = 110 U2 = 101 U6 = 100 U1 = 01 U4 = 00

3) Να συγκριθούν οι κώδικες των ερωτημάτων 1) και 2) ως προς την επίδοσή τους **[1 μονάδα]**. Υπόδειξη: Χρησιμοποιήστε τον τύπο για επίδοση κώδικα : $a = \dfrac{H(C)}{\sum_{i=1}^{n} p_i l_i \ \log_2 q}$

Από την επίλυση των ερωτημάτων a) και b) παρατηρούμε ότι για κάθε ένα από τα σύμβολα U1 … U7 και στους δύο κώδικες τα μήκη li είναι ίδια.
Επειδή και τα υπόλοιπα μεγέθη του τύπου της επίδοσης είναι ίδια, προκύπτει ότι η επίδοση των δύο κωδίκων είναι ίδια.