Data Coding and Compression.

Second Exam

DEEC, IST January 24, 2015

Name: _____ Number: _____

Potentially useful facts: $\log_2 3 \simeq 1.585$; $\log_2 5 \simeq 2.322$; $\log_{10}(2) \simeq 0.30$; $\log_a b = (\log_c b)/(\log_c a)$.

Part I

Exam: questions 1 to 20; correct answer = 0.5, wrong answer = – 0.25.
Test 1: questions 1 to 10; correct answer = 1, wrong answer = – 0.5.
Test 2: questions 11 to 20; correct answer = 1, wrong answer = – 0.5.

1. The random variables $X \in \{1, 2\}$ and $Y \in \{1, 2, ..., 6\}$ represent the outcomes of the throws of a fair coin and a fair die. and $T = X + Y$ denotes their sum; therefore,

   a) $H(T) < \log_2 7$ bits/symbol; ■
   b) $H(T) = \log_2 7$ bits/symbol; □
   c) $H(T) > \log_2 7$ bits/symbol. □

2. Let X, Y, and T be as defined in question 1, and $U = X - Y$ be another random variable; therefore,

   a) $H(U) < H(T)$; □
   b) $H(U) = H(T)$; ■
   c) $H(U) > H(T)$. □

3. Let X, Y, and T be the random variables defined in question 1; therefore,

   a) $I(T;X) < 1$ bits/symbol; ■
   b) $I(T;X) = 1$ bits/symbol; □
   c) $I(T;X) > 1$ bits/symbol. □

4. Let X and Y be the random variables defined in question 1, and $Z = 10\,X + Y$; therefore,

   a) $I(Z;X) < 1$ bits/symbol; □
   b) $I(Z;X) = 1$ bits/symbol; ■
   c) $I(Z;X) > 1$ bits/symbol. □

5. Let X, Y, T, and Z be the random variables defined in questions 1 and 4; therefore,

   a) $I(T;Z) < H(T)$; □
   b) $I(T;Z) = H(T)$; ■
   c) $I(T;Z) > H(T)$. □

6. Let T be the random variable defined in question 1. In an optimal code for variable T,

   a) all codewords have necessarily the same length; □
   b) there are necessarily codewords with different lengths; ■
   c) none of the previous answers. □

7. Consider the source $X \in \{a, b, c, d, e\}$ with probabilities $P(a) = 1/2$, $P(b) = 1/4$, $P(c) = 1/8$, $P(d) = 1/16$ and $P(e) = 1/16$. The expected length of an optimal ternary code for this source is

   a) 5/4 trit/symbol; ■

   b) 6/4 trit/symbol; □

   c) 7/4 trit/symbol □

8. Consider a source generating symbols of the alphabet $\{a, b, c, d\}$ with probabilities $\{1/2 - \varepsilon, 1/4 + \varepsilon, 1/8, 1/8\}$, where $\varepsilon \in [0, 1/2[$, and the code $\{C(a) = 0, C(b) = 11, C(c) = 100, C(d) = 101\}$, which is clearly optimal if $\varepsilon = 0$. The value of ε above which this code is not optimal is

   a) 1/8. ■

   b) 1/4. □

   c) none of the previous values. □

9. Consider a first order Markovian source with the following transition matrix (where $\varepsilon \in [0, 1/8]$)

$P(X_n \mid X_{n-1})$	$X_t = a$	$X_t = b$	$X_t = c$	$X_t = d$
$X_{t-1} = a$	$1/2 - \varepsilon$	0	$1/4 + \varepsilon$	1/4
$X_{t-1} = b$	1/4	$1/2 - \varepsilon$	0	$1/4 + \varepsilon$
$X_{t-1} = c$	$1/4 + \varepsilon$	1/4	$1/2 - \varepsilon$	0
$X_{t-1} = d$	0	$1/4 + \varepsilon$	1/4	$1/2 - \varepsilon$

   The expected length of an optimal binary coding scheme for this source is

   a) $(3/2 - \varepsilon)$ bit/symbol; □

   b) $(3/2 + \varepsilon)$ bit/symbol; ■

   c) 3/2 bit/symbol, for any $\varepsilon \in [0, 1/8]$. □

10. Consider the source described in question 9. The expected length of an optimal ternary coding scheme for this source is

    a) $(1 - \varepsilon)$ trit/symbol; □

    b) $(1 + \varepsilon)$ trit/symbol; □

    c) 1 trit/symbol, for any $\varepsilon \in [0, 1/8]$. ■

11. The Elias Delta code word 0011000001 represents the natural number

    a) 33; ■

    b) 17; □

    c) none of the above. □

12. Which of the following sequences results from the Lempel-Ziv-Welch (LZW) decoding of the sequence *1,5,6,7,8*, (assuming that the alphabet is $\{a, b, c, d\}$ and the first index of the dictionary is 1)?

    a) *aaaaaaaaaaaaaaa* (15 a's); ■

    b) *aaaaaaaaaaaaaaaa* (16 a's); □

    c) *aaaaaaaaaaaaaaaaa* (17 a's). □

13. Consider an LZW coder for the ASCII alphabet $\{..., a, b, ..., A, ..., Z, ...\}$ with 256 symbols (thus represented by 8-bit words) using a dictionary of size $65536 = 2^{16}$ (thus indexed by 16-bit words). The minimum length of a sequence of consecutive b's ("bb...b") such that its LZW compression is not longer than the sequence itself is

    a) 5; □

    b) 6; ■

    c) 7. □

14. Consider a random variable $X \in [0, 1]$, with the probability density function $f_X(x) = 1 + \delta - 2\delta x$, where δ is a parameter in [0, 1]. Then, for any $\delta \in [0, 1]$,

a) $h(X) < 0$; ☐

b) $h(X) > 0$; ☐

c) none of the previous answers. ■

15. Consider the random variable $X \in [0, 1]$ defined in the previous question. Then,

a) $h(X)$ is a monotically increasing function of $\delta \in [0, 1]$; ☐

b) $h(X)$ is a monotically decreasing function of $\delta \in [0, 1]$; ■

c) none of the previous answers. ☐

16. Consider the random variable $X \in [0, 1]$ defined in the previous question, with $\delta = 0$, connected to a non-uniform quantizer with the following 4 regions: $R_0 = [0, 1/2]$, $R_1 = [1/2, 3/4[$, $R_2 = [3/4, 7/8[$, and $R_3 = [7/8, 1]$. The optimal representatives of these regions

a) are located to left of their centers; ☐

b) are located precisely in their centers; ■

c) are located to the right of their centers. ☐

17. Consider the random variable $X \in [0, 1]$ defined in question 14, now with $\delta = 1$, connected to a non-uniform quantizer with the following two regions: $R_0 = [0, 1 - 1/\sqrt{2}]$ and $R_0 =]1 - 1/\sqrt{2}, 1]$. The entropy of the discrete random variable at the output of the quantizer is

a) less than 1 bit/symbol; ☐

b) equal to 1 bit/symbol; ■

c) larger than 1 bit/symbol. ☐

18. Consider the random variable $X \in [0, 1]$ defined in question 14, now with $\delta = 0$, connected to the non-uniform quantizer defined in question 17. The resulting quantization mean squared error (MSE) is

a) less than 1/48; ☐

b) equal to 1/48; ☐

c) larger than 1/48. ■

19. Consider two uniform random variables $X \in [-A, A]$ and $Y \in [-2A, 2A]$, both connected to the corresponding optimal 8-bit quantizers. The mean squared error achieved in quantizing Y is

a) two times larger than that achieved in quantizing X; ☐

b) four times larger than that achieved in quantizing X; ■

c) the same as that achieved in quantizing X. ☐

20. Consider a pair of random variables $(X_1, X_2) \in [-1, 1]^2$, which are independent and uniformly distributed, that is, $f_{X_1,X_2}(x_1, x_2) = 1/4$. The mean squared error (MSE) per component achieved by an optimal 10-bit vector quantizer, compared to that obtained by a pair of 5-bits scalar quantizers (one for each variable),

a) is strictly smaller; ■

b) is the same; ☐

c) none of the previous answers. ☐

Part II

Exam: Problems 1, 2, and 3. Test 1: Problems 1 and 2. Test 2: Problem 3.

Problem 1

Consider the random variable $X \in \{a, b, c, d, e, f\}$ associated with the symbols emitted by a memoryless source, with the following probability distribution: $P(a) = 1/2$, $P(b) = 1/4$, $P(c) = 1/8$, $P(d) = 1/16$, $P(e) = 1/32$, $P(f) = 1/32$.

1. Compute the entropy of the source, $H(X)$, and determine a binary Huffman code. What is the expected length and the efficiency of the obtained code?

2. Determine a ternary Huffman code and compute the resulting expected length. Is this code more or less efficient than the one obtained in the previous question? Why?

3. Consider now that the probability distribution is the following: $P(a) = 1/3$, $P(b) = 1/3$, $P(c) = 1/9$, $P(d) = 1/9$, $P(e) = 1/9$, $P(f) = 0$. Determine a ternary Huffman code and compute the resulting expected length and efficiency. Is this code more or less efficient than the one obtained in the previous question? Why?

4. Consider now that the probability distribution is the following: $P(a) = 1/3$, $P(b) = 1/3 - \varepsilon$, $P(c) = 1/9$, $P(d) = 1/9$, $P(e) = 1/9 + \varepsilon$, $P(f) = 0$, where $\varepsilon \in [0, 1/3[$. Above which value of ε is the code obtained in the previous question not optimal?

Problem 2

1. Consider a first order stationary Markovian source with alphabet $\{a, b, c, d\}$ and the following transition matrix,

$P(X_t \mid X_{t-1})$	$X_t = a$	$X_t = b$	$X_t = c$	$X_t = d$
$X_{t-1} = a$	1/2	1/4	1/8	1/8
$X_{t-1} = b$	1/8	1/2	1/4	1/8
$X_{t-1} = c$	1/8	1/8	1/2	1/4
$X_{t-1} = d$	1/4	1/8	1/8	1/2

   for which the stationary distribution is uniform. Compute the conditional entropy rate of this source.

2. Consider the source in the previous question; determine an optimal binary coding scheme, the corresponding expected code length, and the efficiency.

3. Consider a first order stationary Markovian source with alphabet $\{a, b, c, d\}$ and the following transition matrix,

$P(X_t \mid X_{t-1})$	$X_t = a$	$X_t = b$	$X_t = c$	$X_t = d$
$X_{t-1} = a$	1/3	2/3	0	0
$X_{t-1} = b$	0	1/3	2/3	0
$X_{t-1} = c$	0	0	1/3	2/3
$X_{t-1} = d$	2/3	0	0	1/3

   Determine an optimal binary coding scheme for this source and the respective average codeword length.

4. Consider the second order extension of this source, obtain the corresponding optimal binary coding scheme, compute the resulting expected code length, and compare with the result of the previous question.

Problem 3

Consider a random variable with the following probability density function

$$f_X(x) = \begin{cases} \dfrac{1}{2\,\pi} + \alpha \left(\sin|x| - \dfrac{2}{\pi} \right) & \Leftarrow \quad x \in [-\pi, \pi[, \\ 0 & \Leftarrow \quad x \notin [-\pi, \pi[. \end{cases}$$

where $\alpha \in [-1/4,\, 1/4]$ is a parameter. Notice that $\int_{-\pi}^{0} \left(\sin|x| - \frac{2}{\pi}\right) dx = \int_{0}^{\pi} \left(\sin|x| - \frac{2}{\pi}\right) dx = 0$.

1. Consider $\alpha = 0$ and that X is connected to a non-uniform 2-bit scalar quantizer with the following four regions: $R_0 = [-\pi,\, -3\pi/4]$, $R_1 =]-3\pi/4, -\pi/2]$, $R_2 =]-\pi/2,\, 0]$ e $R_3 =]0,\, \pi]$. Determine the optimal representative of each region and the exact value of the mean squared error of the resulting quantizer.

2. Consider still that $\alpha = 0$ and determine the exact value of the MSE achieved by a uniform 3-bit quantizer.

3. Consider now a generic $\alpha \in [-1/4,\, 1/4]$ and a 1-bit quantizer with regions $R_0 = [-\pi,\, 0[$ and $R_1 = [0,\, \pi]$. Justify that the optimal representatives of these regions, y_0 and y_1, do not depend on α and give their values.

4. Consider still a generic $\alpha \in [-1/4,\, 1/4]$ and a 1-bit optimal quantizer with regions $R_0 = [-\pi,\, 0[$ and $R_1 = [0,\, \pi]$; knowing that

$$\int_{0}^{\pi/2} f_X(x) \left(x - \frac{\pi}{2}\right)^2 dx = \frac{1}{48}\left(-96\,\alpha + \pi^2 + 8\,\alpha\,\pi^2\right),$$

   determine the corresponding exact value of the MSE. Compute the high resolution approximation, compare with the exact value, and comment.

5. Still for a generic $\alpha \in [-1/4,\, 1/4]$, consider now a 2-bit quantizer, with regions $R_0 = [-\pi,\, -\frac{\pi}{2}[$, $R_1 = [-\frac{\pi}{2},\, 0[$, $R_2 = [0,\, \frac{\pi}{2}[$, and $R_3 = [\frac{\pi}{2},\, \pi]$. Knowing that

$$\int_{0}^{\pi/2} x\, f_X(x)\, dx = \frac{1}{16}\left(\pi + \alpha(16 - 4\pi)\right),$$

   find the optimal representatives of the four regions.