

# Data Coding and Compression

Master Program in Electrical and Computer Engineering, IST

January 13, 2017

Name: \_\_\_\_\_

Number: \_\_\_\_\_

- Notes:**
1. Exam (3 hours): everything. Second test (90 minutes): questions 16–30 of Part I and Problem 2.
  2. Part I (exam): correct answer = 0.5 point; wrong answer =  $-0.25$  valores.
  3. Part I (second test): correct answer = 1 valor; wrong answer =  $-0.5$  valores.

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

## Part I

1. Let  $(X_1, X_2, X_3)$  be the random sequence of three cards randomly extracted (**without** replacement) from a standard deck of 52 cards. Then,
  - a)  $H(X_1) < \log_2 52$  bits/symbol;
  - b)  $H(X_1) = \log_2 52$  bits/symbol;
  - c)  $H(X_1) > \log_2 52$  bits/symbol.
2. Let  $(X_1, X_2, X_3)$  be as explained in question 1. Then,
  - a)  $H(X_1, X_2, X_3) < 3 \log_2 52$ ;
  - b)  $H(X_1, X_2, X_3) = 3 \log_2 52$ ;
  - c)  $H(X_1, X_2, X_3) > 3 \log_2 52$ .
3. Let  $(X_1, X_2, X_3)$  be as explained in question 1. Then,
  - a)  $H(X_3|X_2, X_1) < \log_2 52$ ;
  - b)  $H(X_3|X_2, X_1) = \log_2 52$ ;
  - c)  $H(X_3|X_2, X_1) > \log_2 52$ .
4. Let  $(X_1, X_2, X_3)$  be as explained in question 1. Then,
  - a)  $H(X_3|X_2) < H(X_3|X_2, X_1)$ ;
  - b)  $H(X_3|X_2) = H(X_3|X_2, X_1)$ ;
  - c)  $H(X_3|X_2) > H(X_3|X_2, X_1)$ .
5. Let  $(X_1, X_2, X_3)$  be as explained in question 1. Then,
  - a)  $I(X_2; X_1) < \log_2(52/51)$ ;
  - b)  $I(X_2; X_1) = \log_2(52/51)$ ;
  - c)  $I(X_2; X_1) > \log_2(52/51)$ .
6. Let  $(X_1, X_2, X_3)$  be the random sequence of three cards randomly extracted (**with** replacement) from a standard deck of 52 cards. Then,
  - a)  $I(X_2; X_1) < \log_2(52/51)$ ;
  - b)  $I(X_2; X_1) = \log_2(52/51)$ ;
  - c)  $I(X_2; X_1) > \log_2(52/51)$ .
7. Let  $(X_1, X_2, X_3)$  be as explained in question 6. Then,
  - a)  $H(X_3|X_2) < H(X_3|X_2, X_1)$ ;
  - b)  $H(X_3|X_2) = H(X_3|X_2, X_1)$ ;
  - c)  $H(X_3|X_2) > H(X_3|X_2, X_1)$ .

8. Let  $(X_1, X_2, \dots, X_t, \dots, X_{52})$  be the random sequence of obtained by randomly extracting all the cards (**without** replacement) from a standard deck of 52 cards. Then,
- a)  $(X_1, X_2, \dots, X_t, \dots, X_{52})$  is a Markov process of first order;
  - b)  $(X_1, X_2, \dots, X_t, \dots, X_{52})$  is a Markov process of second order;
  - c) none of the previous answers.
9. Let  $(X_1, X_2, \dots, X_t, \dots)$  be the random sequence obtained by randomly extracting cards from a standard deck of 52 cards, **with** replacement of the penultimate card; that is, after extracting  $X_t$ , you put  $X_{t-1}$  back into the deck. Then,
- a)  $(X_1, X_2, \dots, X_t, \dots)$  is a Markov process of first order;
  - b)  $(X_1, X_2, \dots, X_t, \dots)$  is a Markov process of second order;
  - c) none of the previous answers.
10. Consider a discrete memoryless source with alphabet  $\mathcal{X} = \{a, b, c, d\}$ , with probabilities  $p_a = 1/12$ ,  $p_b = 1/4$ , and  $p_c = p_d = 1/3$ ; consider the following two codes:  $\{C_1(a) = 110, C_1(b) = 111, C_1(c) = 10, C_1(d) = 0\}$  and  $\{C_2(a) = 00, C_2(b) = 01, C_2(c) = 10, C_2(d) = 11\}$ .
- a) the code  $C_1$  is optimal and the code  $C_2$  is not optimal;
  - b) the code  $C_2$  is optimal and the code  $C_1$  is not optimal;
  - c) both codes are optimal.
11. Consider the code  $\{C(a) = 00, C(b) = 01, C(c) = 10, C(d) = 11\}$  and two discrete memoryless sources with alphabet  $\mathcal{X} = \{a, b, c, d\}$ : source  $X_1$  has probabilities  $p_a = 1/12$ ,  $p_b = 1/4$ , and  $p_c = p_d = 1/3$ , while source  $X_2$  has probabilities  $p_a = p_b = p_c = p_d = 1/4$ .
- a) The code  $C$  is optimal for source  $X_2$  but not optimal for source  $X_1$
  - b) The code  $C$  is not optimal for any of the two sources;
  - c) The code  $C$  is optimal for both sources.
12. For the source defined in question 10, the expected length of the optimal ternary code is
- a)  $< 4/3$  trits/symbol;
  - b)  $= 4/3$  trits/symbol;
  - c)  $> 4/3$  trits/symbol.
13. Consider a memoryless source  $X \in \{1, 2, 3\}$  with uniform probability distribution. The expected length of an optimal binary code for the second order extension of this source is
- a)  $< 1.6$  bits/symbol;
  - b)  $= 1.6$  bits/symbol;
  - c)  $> 1.6$  bits/symbol.
14. Consider a memoryless source  $Y \in \{1, 2, 3, 4\}$  with uniform probability distribution.
- a) The number of optimal codes for this source is 8.
  - b) The number of optimal codes for this source is 16.
  - c) The number of optimal codes for this source is 24.
15. The efficiency of a code  $C$  for a source  $X$  is defined as  $\eta(C) = \frac{H(X)}{L(C)}$ , where  $L(C)$  is the expected length of  $C$ . Let the sources  $X$  and  $Y$  be as defined in questions 13 and 14, and  $C_X$  and  $C_Y$  be the corresponding optimal binary codes. Then,
- a)  $\eta(C_X) < \eta(C_Y)$ ;
  - b)  $\eta(C_X) = \eta(C_Y)$ ;
  - c)  $\eta(C_X) > \eta(C_Y)$ .

16. The length of the Elias Delta codeword for the natural number 17 is
- a) 9 bits;
  - b) 10 bits;
  - c) none of the previous answers.
17. Assuming that the alphabet is  $\{a, b, c, d, e, f, g, h\}$  and that the dictionary indices start at 0, which of the following sequences results from the Lempel-Ziv-Welch (LZW) coding of the sequence "aaaaaaaaa"?
- a) 0,8,9,9;
  - b) 0,8,9,8;
  - c) none of the previous answers.
18. Assuming that the alphabet is  $\{a, b, c, d, e, f, g, h\}$  and that the dictionary indices start at 0, which of the following sequences symbol results from the Lempel-Ziv-Welch (LZW) decoding of 1,8,9?
- a) *bbbb*
  - b) *bbbbbb*;
  - c) none of the previous answers.
19. Consider an LZW coder for sequences from the alphabet  $\{a, b, c, d, e, f, g, h\}$  with a dictionary with 256 positions. What is the minimum length of a sequence *hhh...h*, such that its LZW encoding is shorter than the original sequence (both measured in bits)?
- a) 13;
  - b) 14;
  - c) none of the previous answers.
20. Consider a Gaussian random variable  $X \in \mathbb{R}$ , with mean 0 and variance  $\sigma^2$ . For which of these two values of the variance is the differential entropy  $h(X)$  positive?
- a) 0.05;
  - b) 0.06;
  - c) none of the previous answers.
21. Consider a random variable  $Y \in [-\gamma, \gamma]$  with uniform probability density function. The minimum value of  $\gamma$  above which  $h(Y)$  is positive is
- a) 1;
  - b) 2;
  - c) none of the previous answers.
22. Consider a Gaussian random variable  $X \in \mathbb{R}$ , with arbitrary mean and variance  $\frac{1}{12}$ , and a random variable  $Z \in [0, 1]$ , with uniform probability density function. Then,
- a)  $h(X) < h(Z)$ ;
  - b)  $h(X) = h(Z)$
  - c)  $h(X) > h(Z)$ .
23. Consider a Gaussian random variable  $X \in \mathbb{R}$ , with zero mean and variance  $\sigma^2$  and the binary random variable  $Y \in \{0, 1\}$ , which is defined as follows: if  $X \leq \gamma$ , then  $Y = 0$ ; if  $X > \gamma$ , then  $Y = 1$ . Then,
- a) the entropy  $H(Y)$  is a monotonically decreasing function of  $\gamma$ ;
  - b) the entropy  $H(Y)$  is a monotonically increasing function of  $\gamma$ ;
  - c) none of the previous answers.

24. Let  $X$  be a random variable with probability density function  $f_X(x) = (2 - |x|)/4$ , for  $x \in [-2, 2]$ , connected to a 1-bit uniform quantizer (with regions  $R_0 = [-2, 0[$  and  $R_1 = [0, 2]$ ). The optimal representatives of the regions are
- a)  $y_0 = -2/3$  and  $y_1 = 2/3$ ; ■
  - b)  $y_0 = -1$  and  $y_1 = 1$ ; □
  - c) none of the previous answers. □
25. Let  $X$  be the random variable defined in question 24, connected to a 2-bit quantizer (with regions  $R_0 = [-2, -1[$ ,  $R_1 = [-1, 0[$ ,  $R_2 = [0, 1[$ , and  $R_3 = [1, 2]$ ). The optimal representatives of the regions are
- a) located to the left of the center of each region; □
  - b) located at uniform distances from each other ( $y_3 - y_2 = y_2 - y_1 = y_1 - y_0$ ); □
  - c) none of the previous answers. ■
26. Let  $Y = Q(X) \in \{y_0, y_1, y_2, y_3\}$  be the random variable corresponding to the output of the quantizer defined in question 25. Then,
- a)  $H(Y) < 3/2$ ; □
  - b)  $H(Y) = 3/2$ ; □
  - c)  $H(Y) > 3/2$ . ■
27. Let  $X$  be a random variable with probability density function  $f_X(x) = |x|/4$ , for  $x \in [-2, 2]$ , connected to a 1-bit quantizer (with regions  $R_0 = [-2, 0[$  and  $R_1 = [0, 2]$ ) and representatives  $y_0 = -1$  and  $y_1 = 1$ . The resulting mean squared error (MSE) is equal to
- a)  $1/3$ ; ■
  - b)  $1/6$ ; □
  - c) none of the previous answers. □
28. Let  $X$  be a random variable with uniform probability density function in  $[-2, 2]$ , connected to a 3-bit uniform quantizer. The resulting mean squared error (MSE) is equal to
- a)  $1/24$ ; □
  - b)  $1/48$ ; ■
  - c) none of the previous answers. □
29. Let  $X \in [-1, 1]$  be a random variable with the following probability density function  $f_X(x) = 3/4$ , if  $|x| < 1/2$ , and  $f_X(x) = 1/4$ , if  $|x| \geq 1/2$ . The optimal representatives of the regions satisfy  $y_0 = -y_1$ , with
- a)  $y_1 = 1/4$ ; □
  - b)  $y_1 = 3/8$ ; ■
  - c) none of the previous answers. □
30. Let  $X$  be a random variable with the following probability density function (where  $\alpha > 0$ )

$$f_X(x) = \frac{\alpha}{1 - e^{-\alpha}} \begin{cases} e^{-\alpha x}, & \text{if } x \in [0, 1] \\ 0 & \text{if } x \notin [0, 1] \end{cases}$$

connected to a quantizer where the first region is  $R_0 = [0, 0.1[$ . The optimal representative  $y_0$  of region  $R_0$

- a) is a monotonically decreasing function of  $\alpha$ ; ■
- b) is a monotonically increasing function of  $\alpha$ ; □
- c) none of the previous answers. □

## Part II

### Problem 1

Consider a discrete memoryless source  $X \in \{1, 2, 3, 4\}$ , with probabilities  $p(1) = 1/2$ ,  $p(2) = 1/4$ ,  $p(3) = p(4) = 1/8$ . Consider  $Y = f(X)$ , where  $f$  is the following function:

$$f(1) = a, \quad f(2) = b, \quad f(3) = c, \quad f(4) = c$$

- Give exact values of  $H(X)$ ,  $H(Y)$ ,  $H(Y|X)$ ,  $H(X|Y)$ ,  $H(X, Y)$ , and  $I(X; Y)$ .
- Obtain an optimal binary code for the second order extension of  $X$  and compute the corresponding expected length.
- Consider a Markov source  $X = (X_1, X_2, \dots, X_t, \dots)$ , where  $X_t \in \{1, 2, 3, 4\}$ , with transition matrix

$$P = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/2 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/4 & 1/2 & 0 & 1/4 \\ 1/4 & 1/4 & 1/2 & 0 \end{bmatrix}.$$

Compute the conditional entropy rate for this source, design an optimal binary coding scheme, and compute the resulting expected code-length. Is it possible to obtain a more efficient coding scheme by using a source extension? Justify your answer.

- Repeat the previous question, for the following transition matrix

$$P = \begin{bmatrix} 0 & 1/4 & 0 & 3/4 \\ 3/4 & 0 & 1/4 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \end{bmatrix}.$$

### Problem 2

Consider a real-valued random variable  $X \in [-1, 1]$ , with probability density function

$$f_X(x) = \begin{cases} 1/2 - \alpha & \Leftarrow |x| < 1/2 \\ 1/2 + \alpha & \Leftarrow |x| \in [1/2, 1], \\ 0 & \Leftarrow x \notin [-1, 1]. \end{cases}$$

where  $\alpha \in [-1/2, 1/2]$  is a parameter.

- Consider  $\alpha = 0$ . For a 2-bit uniform quantizer (four regions of equal length), find the optimal representatives and the resulting MSE.
- For a 1-bit uniform quantizer (two regions of equal length), find the optimal representatives of the two regions, as a function of  $\alpha$ .
- For a 2-bit uniform quantizer, with optimal region representatives, determine the MSE. Is this quantizer optimal for any  $\alpha \in [-1/2, 1/2]$ ? Justify your answer.