Exame de Compressão e Codificação de Dados

Secção de Telecomunicacções – DEEC,
Instituto Superior Técnico

14 de Fevereiro de 2001

Parte I

Esta parte do exame é constituida por 20 perguntas de resposta múltipla. A cotação é a seguinte: resposta certa, 0.5 valores; resposta errada, -0.25 valores.

1. A entropia H de uma fonte sem memória que gera símbolos de um alfabeto de dimensão N, verifica necessariamente

   a) $H < \log N$

   b) $H \geq \log N$

   c) $H \leq \log N$.

2. Sejam X e Y duas variáveis aleatórias; a igualdade $H(X|Y) = H(X)$ significa que

   a) as entropias de X e de Y são iguais,

   b) X e Y são independentes,

   c) X e Y são deterministicamente dependentes.

3. Considere duas variáveis aleatórias X e Y que tomam valores inteiros. Considere a variável aleatória $Z = X + Y$. Qual das seguintes afirmações é verdadeira?

   a) $H(Z) \geq H(X)$, se X e Y forem independentes,

   b) $H(Z) \geq H(X)$, sempre,

   c) $H(Z|Y) > H(X|Y)$.

4. Qual dos seguintes códigos não pode ser óptimo para qualquer distribuição de probabilidades?

   a) $\{1, 01, 00\}$

   b) $\{11, 01, 10, 001\}$

   c) $\{1, 00, 010, 011\}$

5. Considere o código definido por $\{c(A) = 01,\ c(B) = 11,\ c(C) = 00,\ c(D) = 001\}$. Este código

   a) não verifica a desigualdade de Kraft-McMillan,

   b) é instantâneo,

   c) é univocamente descodificável.

6. Considere 9 esferas aparentemente iguais; 8 dessas esferas pesam 1Kg, e uma pesa 1.05Kg. Tem ao seu dispor uma balança de pratos com a qual consegue determinar se os objectos colocados nos pratos pesam o mesmo (a balança fica em equilíbrio) ou se o conteúdo de um dos pratos pesa mais do que o outro. Qual o número de pesagens necessário para identificar a esfera mais pesada?

   a) 2,

   b) 3,

   c) 4.

7. Considere um código **ternário** (isto é, sobre o alfabeto $\{0, 1, 2\}$) instantâneo com 6 palavras e cujas duas palavra mais curtas são "0" e "1".

   a) Há, pelo menos, duas palavras de comprimento 3.

   b) As restantes palavras podem ter todas comprimento 2.

   c) Nenhuma das afirmações anteriores é verdadeira.

8. Considere uma fonte que gera símbolos do alfabeto $\{A, B, C, D, E\}$ com probabilidades, respectivamente, $\{1/2, 1/4, 1/8, 1/16, 1/16\}$, e os códigos $C_1 = \{c(A) = 11,\ c(B) = 10,\ c(C) = 00,\ c(D) = 010,\ c(E) = 011\}$, e $C_2 = \{c(A) = 1,\ c(B) = 00,\ c(C) = 010,\ c(D) = 0110,\ c(E) = 0111\}$.

   a) Nenhum dos dois códigos é óptimo.

   b) Ambos os códigos são óptimos.

   c) Apenas um dos códigos é óptimo.

9. Uma fonte gera símbolos do alfabeto $\mathcal{X} = \{x_1, x_2, ..., x_n\}$, com probabilidades $\{P(x_1), P(x_2), ..., P(x_n)\}$. O código binário óptimo para esta fonte, cujos comprimentos são $l^{\text{opt}}(x_i)$, verifica

   a) $l^{\text{opt}}(x_i) = -\log P(x_i)$, se $P(x_i) = 2^{-l^{\text{opt}}(x_i)}$

   b) $l^{\text{opt}}(x_i) \leq \lceil -\log P(x_i) \rceil$, sempre.

   c) $\sum_i 2^{-l^{\text{opt}}(x_i)} < 1$, sempre.

10. Considere uma fonte que gera símbolos do alfabeto $\{A, B, C\}$ com probabilidades, respectivamente, $\{1/3, 1/3, 1/3\}$. Seja C_1 um código de Huffman para esta fonte, e C_2 um código de Huffman para a extensão de segunda ordem da fonte.

    a) C_2 é mais eficiente do que C_1.

    b) C_1 é mais eficiente do que C_2.

    c) Nenhuma das afirmações anteriores é verdadeira.

11. Considere dois canais binários simétricos, ambos com probabilidade de erro p. A capacidade do canal resultante da ligação em série destes dois canais, designada C_3, é

    a) $C_3 = 1 + p(1-p)\log_2[p(1-p)] + (1 - p(1-p))\log_2[1 - p(1-p)]$,

    b) $C_3 = 1 + 2p(1-p)\log_2[2p(1-p)] + (1 - 2p(1-p))\log_2[1 - 2p(1-p)]$,

    c) nenhuma das expressões anteriores.

12. Considere um canal binário simétrico com probabilidade de erro $p \leq 1/2$, e um canal de apagamento com igual probabilidade de erro, isto é, caracterizado por $P(Y = 0|X = 0) = 1 - p$, $P(Y = *|X = 0) = p$, $P(Y = 1|X = 0) = 0$, $P(Y = 0|X = 1) = 0$, $P(Y = *|X = 1) = p$ e $P(Y = 1|X = 1) = 1 - p$. A capacidade do canal de apagamento é

    a) menor ou igual do que a do canal binário simétrico;

    b) igual a zero quando $p = 1/2$;

    c) maior ou igual do que a do canal binário simétrico.

13. Considere um código linear binário constituido por palavras de comprimento 7, cuja distância mínima é 5 e que, como tal, permite corrigir até dois erros. Ao passar por um canal binário simétrico com probabilidade de erro p, a probabilidade de uma palavra ter erros e ser corrigida é dada por

    a) $(1-p)^7 + 7p(1-p)^6 + 21p^2(1-p)^5$,

    b) $7p(1-p)^6 + 21p^2(1-p)^5$,

    c) nenhuma das expressões anteriores

14. Considere uma fonte ternária de Markov, de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

$p(X_n\|X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$
$X_{n-1} = A$	1/2	1/2	0
$X_{n-1} = B$	0	1/2	1/2
$X_{n-1} = C$	1/2	0	1/2

a) A taxa de entropia condicional desta fonte é 1 bit/símbolo;

b) A entropia da distribuição estacionária desta fonte é 1 bit/símbolo;

c) A taxa de entropia condicional desta fonte é 3/2 bit/símbolo.

15. Considere uma fonte ternária de Markov, de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

$p(X_n\|X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$
$X_{n-1} = A$	1/2	1/4	1/4
$X_{n-1} = B$	1/2	1/4	1/4
$X_{n-1} = C$	1/2	1/4	1/4

a) A taxa de entropia condicional desta fonte é menor do que a entropia da distribuição estacionária.

b) A entropia da distribuição estacionária desta fonte é 2 bit/símbolo.

c) A taxa de entropia condicional desta fonte é igual á entropia da distribuição estacionária.

16. Considere o seguinte procedimento de codificação do tipo Lempel-Ziv: cada subsequência de símbolos (1 símbolo = 1 byte = 8 bits) de comprimento superior ou igual a L que já tenha ocorrido anteriormente é substituida por um par de "ponteiros" com o formato (p_2, p_1), indicando que a ocorrência anterior se verificou na posição p_1, a contar do início, e que o comprimento da subsequência é p_2. O ponteiro p_1 é constituido por 24 bits, enquanto que p_2 é escrito da seguinte forma:

$$1\underbrace{11...1}_{p_2\ 1's}0.$$

A cada símbolo não substituido por "ponteiros" é acrescentado um bit 0 no início para que se possa distinguir dos ponteiros. Qual o valor de L para o qual o sistema apresente maior eficiencia?

a) $L = 3$,

b) $L = 4$,

c) $L = 5$.

17. Considere um quantizador escalar uniforme com apenas 4 células $R_1 = [-2, -1[$, $R_2 = [-1, 0[$, $R_4 = [0, 1[$, $R_4 = [1, 2[$. A entrada deste quantizador é uma variável aleatória X uniformemente distribuida no intervalo $[-2, 2]$. A relação sinal/ruído de quantização ígual a

a) 12,

b) 16,

c) 20.

18. Um quantizador escalar uniforme possui 200 células granulares iguais $R_i = [\frac{2i}{200}, \frac{2i+2}{200}[$, para $i = 0, 1, ..., 199$. A entrada deste quantizador é uma variável aleatória X cuja função densidade de probabilidade é $f_X(x) = 1 - \frac{x}{2}$, para $x \in [0, 2]$. Os elementos do "code-book" óptimo verificam

a) $y_i < \frac{2i+1}{200}$,

b) $y_i = \frac{2i+1}{200}$,

c) $y_i > \frac{2i+1}{200}$.

19. Um quantizador escalar uniforme possui 200 células granulares iguais $R_i = [-1 + \frac{2i}{200},\ -1 + \frac{2i+2}{200}[$, para $i = 0, 1, ..., 199$. Assumindo que a entrada está uniformemente distribuida e que o o "code-book" escolhido para este quantizador é $\{y_i,\ i = 0, 1, ..., 199\}$, com $y_i = -1 + \frac{2i+2}{200}$, o valor quadrático médio do erro de distorção granular é

   a) $10^{-4}/3$,

   b) $10^{-4}/6$,

   c) $10^{-4}/12$.

20. Considere um quantizador vectorial Q_1 aplicado a vectores de dimensão n, e que foi ajustado pelo algoritmo de Lloyd. Considere um quantizador vectorial Q_2, também aplicado a vectores de dimensão n, e que é constituido por n quantizadores escalares, um para cada uma das componente do vector de entrada.

   a) A distorção média de Q_2 não pode ser igual á de Q_1.

   b) A distorção média de Q_1 é sempre melhor ou igual do que a de Q_2.

   c) A distorção média de Q_1 é sempre estritamente melhor do que a de Q_2.

Parte II

Problema 1

a) Considere um quantizador escalar uniforme com apenas 2 células iguais $R_1 = [-1,\ 0[$ e $R_2 = [0,\ 1[$. Considere que a entrada é uma variável aleatória cuja função densidade de probabilidade é

$$f_X(x) = \begin{cases} 1 - |x|, & x \in [-1,\ 1] \\ 0, & x \notin [-1,\ 1] \end{cases}$$

Obtenha expressões exactas para os 2 elementos do "code-book": y_1 e y_2.

b) Nas condições da alínea anterior, calcule o valor exacto do erro quadrático médio de quantização e a relação sinal/ruído de quantização.

c) Repita as duas alíneas anteriores, assumindo agora que a entrada é uniformemente distribuida no intervalo $[-1,\ 1]$.

d) Repita as alíneas a) e b) assumindo agora que a entrada apenas toma os valores -0.25 e 0.25, ambas com probabilidade 1/2, isto é, $f_X(x) = 0.5\,\delta(x - 0.25) + 0.5\,\delta(x + 0.25)$.

Problema 2

a) Considere uma fonte que gera símbolos do alfabeto $\{A, B, C, D, E, F\}$ com probabilidades respectivamente $\{1/3, 1/3, 1/4, 1/24, 1/48, 1/48\}$. Desenhe dois códigos de Huffman com palavras de comprimentos diferentes. Calcule os comprimentos médios desses códigos.

b) Considere uma fonte que gera símbolos do alfabeto $\{A, B, C\}$ com probabilidades respectivamente $\{1/2, 1/4, 1/4\}$. Desenhe um código de Huffman para esta fonte. A saída do codificador pode ser considerada uma nova fonte, agora binária. Calcule as probabilidades de esta fonte emitir o bit 0 e o bit 1, bem como a sua entropia, e comente o resultado.

c) Considere a fonte da alínea anterior e o código $\{c(A) = 01, c(B) = 00, c(C) = 1\}$. A saída deste codificador é uma fonte binária. Calcule as probabilidades de esta fonte emitir o bit 0 e o bit 1, bem como a sua entropia, e comente o resultado.