Exame de Compressão e Codificação de Dados

Secção de Telecomunicacções – DEEC,
Instituto Superior Técnico

19 de Janeiro 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 de uma fonte sem memória que gera símbolos de um alfabeto de dimensão N, com equiprobabilidade, é

   a) proporcional a N,

   b) inversamente proporcional a N,

   c) uma função não linear de N.

2. Sejam X e Y duas variáveis aleatórias; a igualdade $H(X|Y) = H(Y|X)$ verifica-se se

   a) $H(X) = H(Y)$,

   b) só se X e Y forem independentes,

   c) em nenhuma das condições anteriores.

3. Considere a seguinte medida de dependência entre duas variáveis aleatórias X e Y: $\delta(X,Y) = I(X;Y)/H(X)$. Verifica-se que $\delta(X;Y) = 1$ quando

   a) $H(X) = H(Y)$,

   b) $H(X|Y) = 0$,

   c) $H(X|Y) = H(X)$.

4. O código definido por $\{c(A) = 001,\ c(B) = 01,\ c(C) = 10,\ c(D) = 00\}$

   a) é univocamente descodificável,

   b) verifica a desigualdade de Kraft-McMillan, logo é instantâneo,

   c) não verifica a desigualdade de Kraft-McMillan.

5. O código definido por $\{c(A) = 00,\ c(B) = 01,\ c(C) = 10,\ c(D) = 110\}$

   a) não verifica a desigualdade de Kraft-McMillan,

   b) não é óptimo para nenhuma distribuição de probabilidades $\{P(A), P(B), P(C), P(D)\}$,

   c) não é instantâneo.

6. Considere um código binário instantâneo, óptimo para uma determinada fonte, cuja palavra mais comprida tem k bits. Neste caso,

   a) o número de palavras deste código pode ser igual a k;

   b) o número de palavras deste código é maior ou igual a $k+1$;

   c) o número de palavras deste código é necessariamente menor do que 2^k.

7. Considere um código binário instantâneo com 5 palavras e cuja palavra mais curta tem apenas um bit.

   a) As restantes palavras têm no mínimo 3 bits.

   b) A mais longa das restantes palavras tem no mínimo 4 bits.

   c) Nenhuma das afirmações anteriores é verdadeira.

8. Considere uma fonte que gera símbolos do alfabeto $\{A, B, C, D, E\}$ com probabilidades, respectivamente, $\{1/3, 1/3, 1/4, 1/24, 1/24\}$, 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 comprimento ideal de Shannon para os símbolos é $l(x_i) = -\log P(x_i)$. No código óptimo para esta fonte, cujos comprimentos são $l^{\text{opt}}(x_i)$, todos os símbolos x_i verificam

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

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

   c) nenhuma das desigualdades anteriores anteriores é sempre verdadeira.

10. Considere uma fonte que gera símbolos do alfabeto $\{A, B, C, D\}$ com probabilidades, respectivamente, $\{1/2, 1/4, 1/8, 1/8\}$. 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. Sejam C_1 e C_2 as capacidades de dois canais binários, não necessáriamente simétricos. A capacidade do canal resultante da ligação em série dos dois canais, designada C_3, verifica

    a) $C_3 = \min\{C_1, C_2\}$ sempre;

    b) $C_3 < \min\{C_1, C_2\}$ sempre;

    c) nenhuma das afirmações anteriores.

12. Considere um canal binário simétrico com probabilidade de erro p, e um canal "Z" com igual probabilidade de erro, isto é, caracterizado por $P(Y = 0|X = 0) = 1$ e $P(Y = 0|X = 1) = p$. Não é necessário conhecer a expressão para capacidade do canal "Z" para se poder afirmar que a sua capacidade 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 4, cuja distância mínima é 3 e que, como tal, permite corrigir erros simples. Ao passar por um canal binário simétrico com probabilidade de erro 0.1, a probabilidade de uma palavra com erros não ser corrigida é

    a) 0.9477,

    b) 0.2710,

    c) 0.0523.

14. Considere um código ternário (alfabeto $\{0, 1, 2\}$) constituido por palavras de comprimento 4. A distância mínima deste código é 3 e, como tal, permite corrigir erros simples. Ao passar por um canal ternário simétrico com probabiliade de erro 0.1, a probabilidade de uma palavra com erros não ser corrigida é

   a) 0.1808

   b) 0.6144

   c) nenhum dos valores anteriores.

   Sugestão: recorde a expressão do volume de uma esfera de raio r no espaço das palavras de comprimento n num alfabeto q-ário (isto é, o número de palavras a distância menor ou igual a r de uma dada palavra):

$$\sum_{i=0}^{r} \binom{n}{i} (q-1)^i.$$

15. Considere uma fonte binária com memória caracterizada pelas seguintes probabilidades de transição: $P(x_n = 1|x_{n-1} = 0) = \alpha$ e $P(x_n = 0|x_{n-1} = 1) = \beta$. Seja H_1 o valor da taxa de entropia condicional desta fonte. Considere uma outra fonte binária, esta sem memória, caracterizada por $P(x = 0) = \beta/(\alpha + \beta)$, e cuja entropia é H_2. Qual das seguintes afirmações é **falsa**?

   a) $H_2 \geq H_1$;

   b) $H_2 = 1$ bit/símbolo, se $\alpha = 1/2$;

   c) $H_2 = 1$ bit/símbolo, se $\alpha = \beta$;

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_1, p_2), 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:

$$010\underbrace{11...1}_{p_2\ 1's}010.$$

   O valor mínimo de L para o qual este sistema nunca produz uma sequência de bits mais longa do que a original

   a) $L = 4$,

   b) $L = 5$,

   c) $L = 6$.

17. Considere um quantizador escalar uniforme com células granulares de largura Δ. O erro de distorção granular,

   a) tem valor quadrático médio aproximadamente, mas nunca exactamente, igual a $\Delta^2/12$;

   b) tem valor quadrático médio que depende dos valores do "code-book", isto é, dos representates de cada célula;

   c) é sempre uniformemente distribuido no intervalo $[-\Delta/2,\ \Delta/2]$.

18. 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, o "code-book" óptimo para este quantizador é $\{y_i,\ i = 0, 1, ..., 199\}$, com

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

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

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

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 $y_i = -1 + \frac{2i}{200}$, 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 escalar arbitrário. O valor quadrático médio do erro de distorção de sobrecarga

a) é sempre estritamente maior do que zero;

b) pode ser igual a zero;

c) não depende da função densidade de probabilidade da entrada do quantizador.

Parte II

Problema 1

a) Considere um quantizador escalar uniforme com N células iguais no intervalo $[1,\, e]$. Obtenha a expressão que define as células deste quantizador.

b) Assuma que a entrada é uma variável aleatória X cuja função densidade de probabilidade é

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

Obtenha o "code-book" óptimo.

Sugestão: comece por obter o representante óptimo para a célula genérica $R = [a,\, b[$, e depois particularize os valores de a e b utilizando o resultado da alínea (a).

c) Com $N = 100$, obtenha o valor quadrático médio do erro de distorção, granular D_{mse} e a relação sinal/ruído de quantização,

$$\text{SNR} = 10 \log_{10} \frac{\sigma_X^2}{D_{mse}}$$

em que σ_X^2 é a variância da entrada X.

Problema 2

a) Considere um canal cuja entrada é uma variável X que toma valores 0 ou 1 com probabilidades α e $1-\alpha$, respectivamente. A saída é dada por $Y = X + Z$ em que Z é toma os valores 0 ou a com probabilidades p e $1-p$, respectivamente. X e Z são independentes. Desenhe o grafo deste canal e escreva a sua matriz de transição assumindo que $a \neq 1$.

b) Repita a alínea anterior com $a = 1$.

c) Mostre, de preferência sem recorrer a cálculos, que no caso em que $a \neq 1$, a capacidade do canal é igual a 1 bit/símbolo, sendo pois equivalente a um canal binário simétrico com probabilidade de erro nula.

d) Para $a = 1$ e $p = 1/2$, calcule a capacidade do canal resultante e comente o resultado.