

Exame de Compressão e Codificação de Dados

Secção de Telecomunicações – DEEC, **Instituto Superior Técnico**

21 de Janeiro de 2002

Parte I

Cotação da Parte I: 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 ,
 - a) verifica necessariamente $H > 0$;
 - b) verifica necessariamente $H \geq 0$;
 - c) verifica necessariamente $H < \log N$.
2. Sejam X e Y duas variáveis aleatórias; a igualdade $I(X; Y) = H(X)$ implica que
 - a) as entropias de X e de Y são iguais,
 - b) X e Y são independentes,
 - c) nenhuma das afirmações anteriores.
3. Considere a seguinte medida normalizada de dependência entre duas variáveis aleatórias X e Y : $d(X, Y) = 1 - (H(X|Y)/H(X))$. Verifica-se que $d(X; Y) = 1$ quando
 - a) as entropias de X e de Y são iguais,
 - b) X é uma função determinística de Y .
 - c) X e Y são independentes.
4. Considere duas variáveis aleatórias X e Y que tomam valores no conjunto $\{-1, 0, 1\}$, ambas com probabilidades $\{1/3, 1/3, 1/3\}$. Considere a variável aleatória $Z = X + Y$. Então
 - a) $H(Z) = \log 3$; b) $H(Z) = \log 5$; c) $H(Z) < \log 5$.
5. Qual dos seguintes códigos não é instantâneo?
 - a) $\{00, 10, 01, 11\}$; b) $\{0, 01, 110, 111\}$; c) $\{1, 00, 010, 011\}$.
6. Qual dos seguintes códigos é ótimo para uma fonte que emite símbolos do alfabeto $\{A, B, C, D\}$ com probabilidades $3/5, 2/15, 2/15, 2/15$?
 - a) $\{00, 01, 10, 11\}$; b) $\{0, 10, 110, 111\}$; c) nenhum dos códigos anteriores.
7. Qual dos seguintes códigos é ótimo para uma fonte que emite símbolos do alfabeto $\{A, B, C, D\}$ com probabilidades $7/24, 6/24, 6/24, 5/24$?
 - a) $\{00, 01, 10, 11\}$; b) $\{0, 10, 110, 111\}$; c) nenhum dos códigos anteriores.
8. Existem sequências de bits que não correspondem a nenhuma sequência de palavras de código...
 - a) ...apenas quando o código não é instantâneo;
 - b) ...apenas quando o código verifica a desigualdade de Kraft-McMillan estritamente, isto é $\sum_i 2^{-l_i} < 1$;
 - c) ...apenas quando o código verifica a desigualdade de Kraft-McMillan com igualdade, isto é $\sum_i 2^{-l_i} = 1$;

9. Considere um código ternário (isto é, sobre o alfabeto $\{0, 1, 2\}$) instantâneo com 5 palavras e cujas duas palavras mais curtas são “0” e “1”.
- Há, pelo menos, duas palavras de comprimento 3.
 - As restantes palavras podem ter todas comprimento 2.
 - Nenhuma das afirmações anteriores é verdadeira.
10. Considere um código binário instantâneo, ótimo para uma dada fonte, com 5 palavras e cuja palavra mais longa tem 4 bits.
- A mais curta das restantes palavras tem 2 bits.
 - A mais curta das restantes palavras tem 1 bit.
 - Nenhuma das afirmações anteriores é verdadeira.
11. 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 $C_2 = \{c(A) = 1, c(B) = 00, c(C) = 010, c(D) = 0110, c(E) = 0111\}$.
- Nenhum dos dois códigos é ótimo.
 - Ambos os códigos são ótimos.
 - Apenas um dos códigos é ótimo.
12. Uma fonte gera símbolos do alfabeto $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$, com probabilidades $\{P(x_1), P(x_2), \dots, P(x_n)\}$. O código binário ótimo para esta fonte, cujos comprimentos são $l^{\text{opt}}(x_i)$, verifica
- $l^{\text{opt}}(x_i) = -\log P(x_i)$, se $P(x_i) = 2^{-l^{\text{opt}}(x_i)}$
 - $l^{\text{opt}}(x_i) \leq \lceil -\log P(x_i) \rceil$, sempre.
 - $\sum_i 2^{-l^{\text{opt}}(x_i)} < 1$, sempre.
13. Considere uma fonte que gera símbolos do alfabeto $\{A, B, C\}$ com probabilidades, respectivamente, $\{1/2, 1/4, 1/4\}$. 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.
- C_2 é mais eficiente do que C_1 .
 - C_1 é mais eficiente do que C_2 .
 - Nenhuma das afirmações anteriores é verdadeira.
14. Considere uma fonte ternária de Markov, de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

$p(X_n X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$
$X_{n-1} = A$	1/3	1/3	1/3
$X_{n-1} = B$	0	1/2	1/2
$X_{n-1} = C$	1/2	0	1/2

- A taxa de entropia condicional desta fonte é $H = 1$ bit/símbolo;
 - A taxa de entropia condicional desta fonte verifica $H < \log_2 3$ bit/símbolo;
 - A taxa de entropia condicional desta fonte é $H = \log_2 3$ bit/símbolo.
15. Considere uma fonte ternária de Markov, de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

$p(X_n X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$
$X_{n-1} = A$	1/3	1/3	1/3
$X_{n-1} = B$	1/3	1/3	1/3
$X_{n-1} = C$	1/3	1/3	1/3

- a) A taxa de entropia condicional desta fonte é menor do que a entropia da distribuição estacionária.
b) A taxa de entropia condicional desta fonte é igual á entropia da distribuição estacionária.
c) A taxa de entropia condicional desta fonte é igual a 1.5 bits/símbolo.
16. Considere uma fonte ternária de Markov, de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

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

- a) A taxa de entropia condicional desta fonte é igual a $\log_2 3$ bits/símbolo.
b) A taxa de entropia condicional desta fonte é igual á entropia da distribuição estacionária.
c) A entropia da distribuição estacionária desta fonte é igual a $\log_2 3$ bits/símbolo.
17. Considere um quantizador escalar uniforme com apenas 4 células $R_1 = [-2, -1[$, $R_2 = [-1, 0[$, $R_3 = [0, 1[$, $R_4 = [1, 2[$. Os representantes de cada célula são, respectivamente, $y_1 = -2$, $y_2 = -1$, $y_3 = 0$, $y_4 = 1$. O erro quadrático médio de quantização deste quantizador, quando a entrada é uma variável aleatória X uniformemente distribuída no intervalo $[-2, 2]$ é igual a
- a) $1/6$,
b) $1/3$,
c) $1/12$.
18. Um quantizador escalar uniforme possui 200 células granulares iguais $R_i = [\frac{2i}{200}, \frac{2i+2}{200}[$, para $i = 0, 1, \dots, 199$. A entrada deste quantizador é uma variável aleatória X cuja função densidade de probabilidade é $f_X(x) = \frac{x}{2}$, 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. Considere uma variável aleatória X cuja função densidade de probabilidade é Gaussiana de média nula e variância σ^2 , isto é, $f_X(x) = (2\pi\sigma^2)^{-1/2} \exp\{-x^2/(2\sigma^2)\}$. Usando um quantizador escalar uniforme com 256 células distribuídas no intervalo $[-\sigma; \sigma]$ o erro quadrático médio de sobrecarga é
- a) nulo,
b) não nulo,
c) depende de σ .
20. Considere o algoritmo de Linde, Buzo e Gray (LBG) para o desenho de quantizadores vectoriais com base num conjunto de amostras da variável aleatória a quantizar. Para um tamanho de “codebook” fixo, o peso computacional deste algoritmo
- a) Não depende do número de amostras.
b) Cresce quadraticamente com o número de amostras.
c) Cresce linearmente com o número de amostras..

Parte II

Problema 1

Uma das importantes contribuições da teoria da informação para a estatística é o *método dos tipos*. Considere uma sequência $\mathbf{x} = x(1), x(2), \dots, x(n)$, de comprimento n , gerada por uma fonte cujo alfabeto é $\mathcal{X} = \{x_1, \dots, x_N\}$. Por exemplo, $\mathbf{x} = ABAABCCABCAB$ é uma sequência de comprimento $n = 12$, gerada por uma fonte de alfabeto $\mathcal{X} = \{A, B, C\}$. O *tipo* da sequência $\mathbf{x}$, designado $P_{\mathbf{x}} = [p_1, \dots, p_N]$, é um vector de frequências relativas de ocorrências de cada símbolo do alfabeto na sequência. Por exemplo, para $\mathbf{x} = ABAABCCABCAB$, temos $P_{\mathbf{x}} = [5/12, 4/12, 3/12]$, correspondente a 5 *A*'s, 4 *B*'s e 3 *C*'s, num total de 12. Seja $Q = [q_1, \dots, q_N]$ a distribuição de probabilidade da fonte. Por exemplo, para uma fonte de alfabeto $\mathcal{X} = \{A, B, C\}$, podemos ter $Q = [0.5, 0.3, 0.2]$, correspondente a $p(A) = 0.5$, $P(B) = 0.3$ e $P(C) = 0.2$.

É possível demonstrar que a probabilidade de uma fonte cuja distribuição de probabilidade é Q gerar uma sequência $\mathbf{x}$ é dada por

$$P(\mathbf{x}) = 2^{-n(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||Q))} \quad (1)$$

onde $H(P_{\mathbf{x}})$ é a entropia do tipo $P_{\mathbf{x}}$ e $D(P_{\mathbf{x}}||Q)$ é a divergência de Kullback-Leibler entre o tipo $P_{\mathbf{x}}$ e a distribuição de probabilidade Q , isto é,

$$D(P_{\mathbf{x}}||Q) = \sum_{i=1}^N p_i \log \frac{p_i}{q_i}.$$

- Obtenha o caso particular da expressão (1) para uma fonte de símbolos equiprováveis, $Q = [1/N, \dots, 1/N]$. Interprete o resultado.
- Obtenha o caso particular da expressão (1) para $P_{\mathbf{x}} = Q$. Aplique a expressão obtida para calcular a probabilidade de uma fonte de alfabeto $\{A, B, C, D\}$, com distribuição de probabilidade $[1/2, 1/4, 1/8, 1/8]$, gerar uma dada sequência de comprimento 40 com exactamente 20 *A*'s, 10 *B*'s, 5 *C*'s e 5 *D*'s.
- Designa-se por $T_n(P_{\mathbf{x}})$ o conjunto de todas as sequências de comprimento n cujo tipo é $P_{\mathbf{x}}$. Por exemplo, para uma fonte binária de alfabeto $\{A, B\}$, $T_4([3/4, 1/4])$ é o conjunto de todas as sequências com 3 *A*'s e um *B*, isto é, $T_4([3/4, 1/4]) = \{AAAB, AABA, ABAA, BAAA\}$. Demonstra-se que a probabilidade de $T_n(P_{\mathbf{x}})$, quando n é grande, é aproximadamente dada por

$$P(T_n(P_{\mathbf{x}})) \simeq 2^{-nD(P_{\mathbf{x}}||Q)}.$$

Com base neste resultado calcule

$$\lim_{n \rightarrow \infty} P(T_n(P_{\mathbf{x}})),$$

para os casos $P_{\mathbf{x}} = Q$ e $P_{\mathbf{x}} \neq Q$, e comente o resultado.

Problema 2

- 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.
- Considere um código de comprimento fixo para a fonte definida em **a)**. Calcule o seu comprimento médio e comente o resultado.
- Considere uma fonte que gera símbolos do alfabeto $\{A, B, C, D, E\}$ com probabilidades respectivamente $\{1/2, 1/8, 1/8, 1/8, 1/8\}$. Calcule o comprimento médio do código óptimo para esta fonte sem usar o procedimento de Huffman.