

Exame de Compressão e Codificação de Dados

Secção de Telecomunicações – DEEC, Instituto Superior Técnico

5 de Fevereiro de 2003

Parte I

Cotação da Parte I: resposta certa, 0.5 valores; resposta errada, -0.25 valores.

Salvo explicitamente indicado, todos os logaritmos são em base 2 e as entropias e comprimentos médios de códigos são expressos em bit/símbolo.

1. Considere o par de variáveis aleatórias (v.a.) X e Y , ambas binárias, cuja função de probabilidade conjunta é dada por: $p(X = 0, Y = 0) = 1/3$, $p(X = 0, Y = 1) = 1/3$, $p(X = 1, Y = 0) = 1/3$, e $p(X = 1, Y = 1) = 0$. Qual das seguintes afirmações é verdadeira?
a) $H(X, Y) = H(X) + H(Y)$; **b)** $H(X) = \log 3$; **c)** $H(Y) = \log 3 - 2/3$.
2. Sejam X e Y as v.a. da pergunta anterior. Então,
a) $H(X|Y) = 2/3$; **b)** $H(X|Y) = \log 2$; **c)** $H(X|Y) = \log 3$.
3. Considere ainda as v.a. da pergunta 1. Então,
a) $I(X; Y) = \log 2$; **b)** $I(X; Y) = \log 3 - 2/3$; **c)** $I(X; Y) = \log 3 - 4/3$.
4. Considere ainda as v.a. da pergunta 1, e uma terceira variável aleatória $Z = Y + 4 \times X$. Então,
a) $H(Z) = \log 3$; **b)** $H(Z) = \log 4$; **c)** Nenhum dos valores anteriores.
5. Considere que X representa uma carta retirada ao acaso de um baralho de 52 cartas. Seja Y o naipe da carta retirada, isto é, $Y \in \{\heartsuit, \spadesuit, \diamondsuit, \clubsuit\}$. Então,
a) $I(X; Y) = \log 13$; **b)** $I(X; Y) = \log 4$; **c)** nenhum dos valores anteriores.
6. Considere que X representa uma carta retirada ao acaso de um baralho de 52 cartas. Seja Y o naipe da carta retirada, isto é, $Y \in \{\heartsuit, \spadesuit, \diamondsuit, \clubsuit\}$, e Z o valor da carta, isto é $Z \in \{\text{“Ás”}, \text{“2”}, \text{“3”}, \dots, \text{“valete”}, \text{“rei”}\}$. Então,
a) $I(Y; Z) = \log 13$; **b)** $I(Y; Z) = \log 4$; **c)** nenhum dos valores anteriores.
7. Uma fonte gera símbolos do alfabeto $\{A, B, C, D, E\}$, com probabilidades $(1/3, 1/3, 1/9, 1/9, 1/9)$. Qual dos seguintes códigos binários é óptimo para esta fonte?
a) $\{1, 01, 000, 0011, 0010\}$; **b)** $\{00, 01, 100, 101, 11\}$; **c)** ambos.
8. A eficiência ($\varepsilon = H(X)/E[L(X)]$) do código binário óptimo para a fonte da questão 7 verifica
a) $\varepsilon = 1$; **b)** $\varepsilon < 1$; **c)** $\varepsilon > 1$.
9. A eficiência ($\varepsilon = H(X)/E[L(X)]$) do código ternário óptimo para a fonte da questão 7 verifica
a) $\varepsilon = 1$; **b)** $\varepsilon < 1$; **c)** $\varepsilon > 1$.
10. Considere uma fonte que gera símbolos de um alfabeto $\{A, B, C\}$ com probabilidades $\{0.6, 0.2, 0.2\}$. Quantos códigos binários óptimos existem para esta fonte?
a) 4; **b)** 6; **c)** 8.
11. Para uma fonte binária com entropia 0.25 bits/símbolo, qual a menor ordem de extensão a partir da qual se garante uma eficiência de codificação superior a 50%?
a) 2; **b)** 3; **c)** 4.

12. Considere uma fonte quaternária (alfabeto $\{A, B, C, D\}$) com entropia 2 bits/símbolo. Qual o comprimento médio do código $\{0, 11, 100, 101\}$, quando utilizado para codificar símbolos dessa fonte?
- a) 2; b) 2.25; c) 2.5.

13. Considere a fonte ternária (alfabeto $\{A, B, C\}$), Markoviana 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
$X_{n-1} = B$	$1/2$	0	$1/2$
$X_{n-1} = C$	$1/3$	$1/3$	$1/3$

Qual das seguintes é a distribuição estacionária desta fonte?

- a) (0.4, 0.3, 0.3); b) (0.3, 0.4, 0.3); c) (0.3, 0.3, 0.4).
14. Considere a fonte quaternária (alfabeto $\{A, B, C, D\}$), Markoviana de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

$p(X_n X_{n-1})$	$X_n = A$	$X_n = B$	$X_n = C$	$X_n = D$
$X_{n-1} = A$	$1/2$	$1/4$	0	$1/4$
$X_{n-1} = B$	$1/4$	$1/2$	$1/4$	0
$X_{n-1} = C$	0	$1/4$	$1/2$	$1/4$
$X_{n-1} = D$	$1/4$	0	$1/4$	$1/2$

A taxa de entropia condicional desta fonte é igual

- a) 1; b) $3/2$; c) 2.
15. Considere a fonte binária (alfabeto $\{A, B\}$), Markoviana de primeira ordem, caracterizada pelas seguintes probabilidades de transição:

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

O comprimento médio do código binário ótimo para esta fonte é

- a) 1; b) $1/2$; c) $2/3$.
16. Um determinado quantizador escalar uniforme com b bits, em aproximação de alta resolução, apresenta um erro quadrático médio de distorção igual a 10. Ao passar o número de bits para $b-1$, o erro quadrático médio passa a ser

- a) 20; b) 40; c) 60.
17. Considere uma fonte X cuja função densidade de probabilidade é $f_X(x) = (1 - |x|)$, para $|x| \leq 1$ aplicada na entrada de um quantizador cujo *code-book* é $\{y_1 = -0.5, y_2 = 0.75\}$. Para este *code-book*, as células ótimas (isto é, aquelas para as quais se obtém o menor erro quadrático médio) são
- a) $R_1 = [-1, 0]$ e $R_2 = [0, 1]$; b) $R_1 = [-1, -0.125]$ e $R_2 = [-0.125, 1]$;
c) $R_1 = [-1, 0.125]$ e $R_2 = [0.125, 1]$.

18. Considere uma variável aleatória X cuja função densidade de probabilidade é $f_X(x) = (1 - |x|)$, para $|x| \leq 1$, ligada na entrada de um quantizador uniforme com 1024 células. Seja $\mathcal{E}(x) \in \{0, 1, 2, \dots, 1024\}$ a função que indica qual o índice da célula a que pertence x . A entropia da v.a. $\mathcal{E}(X)$ verifica

- a) $H(\mathcal{E}(X)) = 10$; b) $H(\mathcal{E}(X)) > 10$; c) $H(\mathcal{E}(X)) < 10$.

19. Qual a principal vantagem dos métodos de codificação de tipo Lempel-Ziv sobre a codificação aritmética?
- a)** Permitem obter maiores taxas de compressão;
 - b)** não requerem uma caracterização probabilística da fonte;
 - c)** são mais rápidos.
20. Considere que se pretende quantizar uma v.a. distribuída no intervalo $[-1, 1]$, de variância unitária. Usando a aproximação de alta resolução, qual o menor número de bits que garante uma relação sinal-ruído de quantização ($SNR_Q = 10 \log_{10}(\sigma_x^2/D_{MSE})\text{dB}$) maior do que 32dB?
- a)** 5; **b)** 6; **c)** 7.

Nota: se necessitar, use as seguintes aproximações: $\log_{10} 3 \simeq 0.5$; $\log_{10} 2 \simeq 0.3$.

Parte II

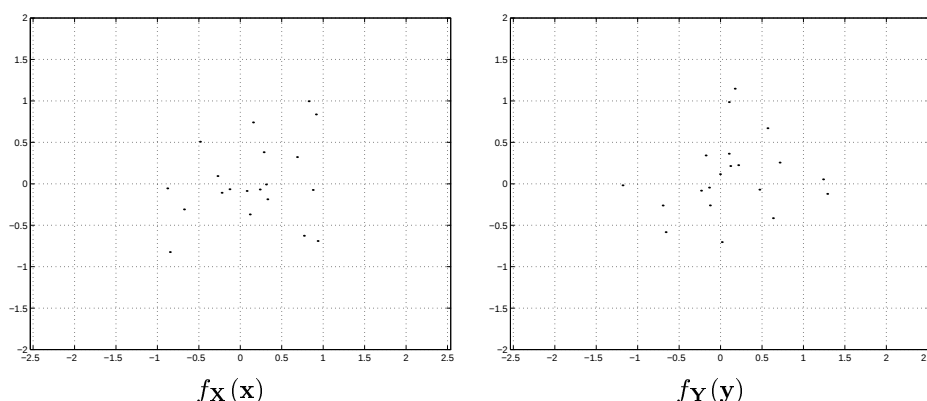
Problema 1 Considere uma variável aleatória X cuja função densidade de probabilidade é $f_X(x) = (1 - |x|)$, para $|x| \leq 1$, aplicada na entrada de um quantizador de 4 células $R_1 = [-1, -1/3]$, $R_2 =]-1/3, 0]$, $R_3 =]0, 1/3]$ e $R_4 =]1/3, 1]$.

- a) Indique as expressões que conduzem ao *code-book* óptimo para estas células.
- b) Qual o valor do erro quadrático médio previsto pela aproximação de alta resolução para células desiguais ($D_{MSE} \simeq \sum_i p_i \Delta_i^2/12$)?
- c) Usando a aproximação de alta resolução para células desiguais ($D_{MSE} \simeq \sum_i p_i \Delta_i^2/12$), obtenha a localização óptima das células. Sugestão: justifique que as células óptimas se podem escrever $R_1 = [-1, -\delta]$, $R_2 =]-\delta, 0]$, $R_3 =]0, \delta]$ e $R_4 =]\delta, 1]$ e tente minimizar D_{MSE} em ordem a δ .

Problema 2 Considere a variável aleatória X que consiste no número de vezes que tenho de lançar uma moeda (equilibrada) até sair uma “cara” pela primeira vez.

- a) Obtenha a expressão geral de $p(X = i)$, para $i = 1, 2, 3, \dots$, e o valor da entropia de X . Expressões que podem úteis: $\sum_{i=1}^{\infty} a^i = a/(1 - a)$; $\sum_{i=1}^{\infty} i a^i = a/(a - 1)^2$.
- b) Desenhe um código instantâneo para codificar X , cujo comprimento médio é precisamente igual à entropia.

Problema 3 Considere um vector aleatório $\mathbf{X} = [X_1, X_2]$, em que X_1 e X_2 são duas variáveis aleatórias independentes e uniformemente distribuídas em $[-1, 1]$. Na figura (à esquerda) ilustra-se a função densidade de probabilidade $f_{\mathbf{X}}(\mathbf{x})$ por meio de 2000 amostras.



- a) Desenhe o quantizador vectorial óptimo de 16 células (4 bits) para $\mathbf{X}$ (basta esboçar, indicando claramente a localização das fronteiras das regiões e dos vectores do *code-book*).
- b) Desenhe quantizadores escalares óptimos, separadamente para X_1 e X_2 , usando 2 bits para cada um. Compare com o resultado da alínea a) e comente.
- c) Considere agora o vector aleatório $\mathbf{Y} = [Y_1, Y_2]$, em que $Y_1 = (X_1 + X_2)/\sqrt{2}$ e $Y_2 = (X_1 - X_2)/\sqrt{2}$. Na figura (à direita) ilustra-se a função densidade de probabilidade $f_{\mathbf{Y}}(\mathbf{y})$ por meio de 2000 amostras. Desenhe o quantizador vectorial óptimo de 16 células (4 bits) para $\mathbf{Y}$.
- d) Explique como podia quantizar de forma óptima o vector aleatório $\mathbf{Y}$, sem recorrer ao quantizador vectorial desenhado em c), mas sim a um par de quantizadores escalares.