# Exame de Compressão e Codificação de Dados

Secção de Telecomunicacções – DEEC
**Instituto Superior Técnico**

25 de Fevereiro de 2000

1. **a)** Considere uma fonte sem memória que emite símbolos do alfabeto $\mathcal{X} = \{A, B, C\}$, com probabilidades $\{p(A) = 0.9, p(B) = 0.05, p(C) = 0.05\}$. Desenhe um código óptimo (por exemplo, de Huffman) para esta fonte e calcule a sua eficiência.

   **b)** Qual a ordem da extensão necessária para que o comprimento médio de codificação seja inferior a 0.7 bits/símbolo ?

   **c)** Uma fonte $S$ gera aleatoriamente símbolos do conjunto com 48 elementos, $\{A, B, C, ..., Z, a, b, ..., z\}$ (todas as letras, maiúsculas e minúsculas). Considere agora que a saída desta fonte é ligada a um bloco que converte todas as letras em minúsculas (por exemplo, "AeCxuDBaY" é transformado em "aecxudbay"). A saída desse bloco de conversão constitui uma outra fonte $X$. Sabendo que a entropia de $S$ é $\log_2 48$ bits/símbolo, determine a entropia de $X$ e a informação mútua $I(S;X)$.

   Considere que a fonte $S$ é ligada a um bloco diferente que converte todas as maiúsculas em minúsculas e as minúsculas em maiúsculas (por exemplo, "AeCxuDBaY" é transformado em "aEcXUdbAy"); designe a saída deste novo bloco por $Y$. Determine a entropia de $Y$ e a informação mútua $I(S;X)$.

   **d)** Existe um procedimento de codificação (proposto por Tunstall, em 1968) que pode ser visto como recíproco do de Huffman. No código de Huffman procura-se codificar os símbolo da fonte com sequências de comprimento variável por forma a que aos símbolos mais prováveis correspondam sequências mais curtas. Na codificação de Tunstall, o objectivo é codificar sequências (com comprimento variável) de símbolos da fonte com palavras de código de comprimento fixo. Ao desenhar um código de Tunstall, o objectivo natural é maximizar o comprimento médio das sequências de entrada do código. Um código de Tunstall de comprimento $n = 2$ para uma fonte de símbolos $\{A, B\}$ com $p(A) = 0.8$ é

   $$AAA \to 00, \;\; AAB \to 01, \;\; AB \to 10, \;\; B \to 11$$

   Calcule o eficiência média deste código. Desenhe um código de Huffman para a extensão de ordem 2 da fonte (isto é, para $\{AA, AB, BA, BB\}$) e compare os resultados.

2. Um instrumento de medida produz valores inteiros no conjunto {0,1,...,511}; esta "fonte" é descrita por um modelo de Markov de primeira ordem dado por

   $$\begin{aligned} p(X_n = X_{n-1} | X_{n-1}) &= 1/2 \\ p(X_n = X_{n-1} - 1 | X_{n-1}) &= 1/4 \\ p(X_n = X_{n-1} + 1 | X_{n-1}) &= 1/4 \end{aligned} \tag{1}$$

   (em que as operações "+" e "-" são tomadas módulo 512, isto é $511 + 1 = 0$ e $0 - 1 = 511$).

   **a)** Verifique que a distribuição estacionária desta fonte é

   $$\mu = [\underbrace{1/512, 1/512, ...., 1/512}_{512 \text{ elementos}}].$$

**b)** Calcule a taxa de entropia condicional, bem como a entropia da distribuição estacionária. Compare e comente.

**c)** Desenhe um sistema de codificação preditiva para esta fonte. Calcule o comprimento médio do código obtido e compare com a taxa de entropia.

3. **a)** Considere um canal binário simétrico (CBS) cuja probabilidade de erro é $\delta$. Qual a capacidade deste canal?

**b)** Considere dois canais binários simétricos iguais, ambos com probabilidade de erro $\delta$, ligados em série. Qual capacidade do canal resultante?

**c)** Considere de novo o CBS com probabilidade de erro $\delta$. Considere que se utiliza um código de repetição com palavras de 3 bits; isto é, para enviar 0 transmite-se 000, para enviar 1, transmite-se 111. O critério de descodificação é o de distância mínima; isto é, por exemplo, 001 é descodificado como 000, e 110 como 111. Mostre que a probabilidade de errar é igual a $3\delta^2 - 2\delta^3$.

**d)** Nas condições da alínea anterior, determine quais os valores de $\delta$ para o quais o código não apresenta nenhuma vantagem (isto é, para os quais $\delta = 3\delta^2 - 2\delta^3$); calcule a capacidade de canal correspondente e interprete o resultado.

4. Considere um quantizador escalar uniforme, com 20 células, cuja gama de valores de entrada é $[-1, 1[$. Assuma que a entrada é uma variável aleatória $X$ cuja função densidade de probabilidade é a que se apresenta na figura seguinte:

**a)** Considerando que $A = 0$, qual o valor da relação sinal/ruído de quantização $SNR_Q$ (em dB)? Recorde que $SNR_Q = 10\log_{10}(P_s^2/P_r^2)\,dB$, em que $P_s^2$ e $P_r^2$ são, respectivamente, os valor médios do quadrado do sinal e do erro de quantização (as potências médias). Este quantizador é óptimo para este caso (com $A = 0$)?

**b)** Recorde que o valor médio do quadrado do erro de quantização, para quantizadores não uniformes, sob a aproximação de alta resolução, é dado por

$$D = \frac{1}{12}\sum_i p_i \Delta_i^2,$$

em que a soma varre todas as células, $p_i$ é a probabilidade da entrada cair na célula $i$, e $\Delta_i$ é a largura da célula $i$. Com $A \neq 0$, o quantizador uniforme ainda é óptimo? Justifique.

**c)** Considere, finalmente, um quantizador ainda com 20 células, das quais 15 estão no intervalo $[-1, 0]$ (portanto cada uma com largura $1/15$) e as restantes 5 no intervalo $[0, 1]$ (portanto cada uma com largura $1/5$). Determine o valor de $A$ a partir do qual este quantizador não uniforme apresenta menor erro de quantização quadrático médio do que o quantizador uniforme original.