Departamento de Informática
Escola de Engenharia
Universidade do Minho

MIEI - Comunicação de Dados
Ficha de Exercícios – Teoria da Informação
Duas aulas

1. No contexto da Teoria da Informação, explique e distinga os seguintes conceitos teóricos: *i)* Informação própria *ii)* Entropia e *iii)* Débito de informação.

2. Uma carta é tirada de um baralho de cartas de jogo.
    a) É informado que a carta que tirou é uma espada. Quanta informação recebeu?
    b) Quanta informação recebe se lhe for dito que a carta que tirou é um ás?
    c) Quanta informação recebe se lhe for dito que a carta que tirou é um ás de espadas? Verifique a relação que existe entre este resultado e os obtidos em a) e b).

3. Calcular o débito de informação de uma fonte telegráfica que emite pontos e traços com probabilidades de ocorrência do ponto e do traço respetivamente P_p=2/3, P_t=1/3, tendo em conta que a fonte emite, em média, 3.75 símbolos por segundo.

4. Uma fonte emite *n* mensagens distintas $\{x_1, ..., x_n\}$ com probabilidades associadas $\{p_1, ..., p_n\}$. Considere o caso em que todas mensagens ocorrem com a mesma probabilidade, i.e. p_i=1/n. Calcule o valor da entropia da fonte e discuta o resultado obtido.

5. Uma fonte emite oito símbolos distintos {A,B,C,D,E,F,G,H} com as seguintes probabilidades: P(A)=1/2, P(B)=P(C)=P(D)=1/12, P(E)= P(F)=P(G)=P(H)=1/16.
    a) Calcule o valor da entropia desta fonte.
    b) Qual o rendimento obtido se na codificação da fonte se utilizar um código de comprimento fixo mínimo.
    c) Codifique a fonte utilizando códigos de *Shannon-Fano* (página 208 da sebenta) e refira qual o rendimento e compressão obtidos.
    d) Indique de que forma poderia ainda tentar melhorar a codificação desta fonte.

6.

	Uma fonte emite oito símbolos distintos {A,B,C,D,E,F,G,H} com as seguintes probabilidades: P(A)=1/2, P(B)=P(C)=P(D)=1/12, P(E)= P(F)=P(G)=P(H)=1/16.
A1	O valor da entropia da fonte (bits/símbolo) poderia ser superior a 3 bits/símbolo caso se assumisse outros valores para as probabilidades dos símbolos.
B2	O valor da entropia desta fonte é superior a 2 bits/símbolo.
C3	Utilizando códigos de *Shannon-Fano*, a transmissão de uma qualquer mensagem com Z símbolos desta fonte requer sempre um número total de dígitos binários inferior a Z*3.
D4	Com codificação por blocos de K símbolos, era possível encontrar um valor de K de tal forma que o comprimento médio do código ($\overline{N}$) fosse inferior a 2.2 dígitos binários por símbolo.

Indique se considera cada uma das afirmações anteriores verdadeira (**V**) ou Falsa (**F**):

A1		**B2**		**C3**		**D4**	

7.

	Uma fonte de informação emite dezasseis símbolos independentes entre si de um alfabeto X, gerando em média 4800 símbolos cada 30 segundos. Sabe-se que o débito de informação desta fonte é de 240 bits/seg.
A1	Com os dados apresentados podemos afirmar que os dezasseis símbolos gerados pela fonte não são equiprováveis.
B2	Com codificação da fonte seria possível obter uma compressão superior a 60%.
C3	Usando códigos binários de comprimento fixo mínimo, para uma codificação por blocos de 3 símbolos (K=3) necessitávamos de um código com comprimento de 12 dígitos binários por cada conjunto de três símbolos.
D4	É possível definir uma codificação binária por blocos de quatro símbolos que permita a obtenção de um comprimento médio de código inferior a 8 dígitos binários por cada conjunto de quatro símbolos.

Indique se considera cada uma das afirmações anteriores verdadeira (**V**) ou Falsa (**F**):

A1		**B2**		**C3**		**D4**	

8. Uma fonte emite quatro símbolos distintos {A,B,C,D} com as seguintes probabilidades P(A)=0.4, P(B)=0.4, P(C)=0.1 e P(D)=0.1.

    a) Determine o valor da entropia da fonte.
    b) Determine um código de comprimento variável para a fonte que possua um rendimento não inferior a 97%.

9. Uma fonte de dados binária produz símbolos 0 e 1 com P_0=3/8 P_1=5/8 e a influência entre símbolos em grupos de dois símbolos sucessivos é tal que $P_{0|1}$=1/16 e $P_{1|0}$=3/4.

    a) Calcule a entropia desta fonte com memória.
    b) Compare o valor obtido em a) com o valor da entropia se a fonte fosse considerada sem memória.
    c) Determine um código de comprimento variável para a fonte considerando blocos de dois símbolos (K=2) e calcule o seu rendimento.

10. Comente a seguinte afirmação: *"Através da utilização de codificações Shannon-Fano é sempre possível obter um código de rendimento superior ao obtido por um código de comprimento fixo mínimo".*

11. Suponha que pretende desenvolver uma aplicação de compressão/descompressão de ficheiros tendo como base a utilização de códigos *Shannon-Fano*. Neste contexto, raciocine sobre os seguintes aspectos:

    - Qual seria a estrutura geral da aplicação a desenvolver e que algoritmos implementaria para as tarefas de compressão/descompressão dos ficheiros?
    - Qual seria a estrutura de um ficheiro comprimido pela sua aplicação?
    - Seria possível que, após utilizar a sua aplicação para compressão de um determinado ficheiro, o ficheiro resultante fosse maior que o ficheiro original?
    - **Sugestão**:
      Implemente um protótipo de uma aplicação deste tipo utilizando uma linguagem de programação da sua preferência. Verifique os níveis de compressão que consegue obter com essa aplicação.

Informação própria do símbolo *S* com probabilidade P_S em fontes sem memória: $I_S = \log_2 P_S^{-1}$ bits

A Entropia de fonte sem memória com *m* símbolos S_i com probabilidades P_i: $H_S = \sum_{i=1}^{m} P_i I_i$ bits/símb

Débito de Informação duma fonte com débito de símbolos r_S: $R_S = r_S H_S$ bits/seg

Comprimento médio dum código com *m* símbolos, cada símbolo com comprimento de N_i bits:
$\overline{N} = \sum_{i=1}^{m} P_i N_i$ bits/símb

Comprimento médio dum Código de Comprimento Fixo Mínimo (CCFM) com *m* símbolos N_f:
$N_f = \log_2(m_{int})$ bits/símb, em que m_{int} é a primeira potência de 2 igual ou maior que *m*

Rendimento dum código: $\rho = H_S/\overline{N}$

Compressão atingida por um código: $c = (N_f - \overline{N})/N_f$ e $c_{max} = (N_f - H_S)/N_f$

Equação que define os limites ao comprimento médio dum código (K é número de blocos na codificação):

$$H_S \leq \overline{N} < H_S + \frac{1}{K}$$

Informação própria condicional dum símbolo *S* em fontes com memória de 1ª ordem e *m* símbolos:
$I_c(S) = \sum_{i=1}^{m} \left(P_{i|S} \log_2 P_{i|S}^{-1} \right)$ bits/símb, em que $P_{i|S}$ é a probabilidade do símbolo S_i aparecer depois do símbolo *S*. Ou seja, a informação própria condicional dum símbolo *S* tem em consideração as diversas probabilidades associadas ao aparecimento desse símbolo antes dum qualquer outro símbolo do alfabeto da fonte, incluindo ele próprio. Esta informação própria também é designada por Entropia Condicional relativa ao símbolo *S*, ou $H_c(S)$.

A Entropia (ou entropia real) em fontes com memória de 1ª ordem e *m* símbolos é dada por:
$H_R = \sum_{i=1}^{m} P_i I_c(S_i)$ bits/símb, em que P_i é a probabilidade simples do símbolo S_i aparecer.