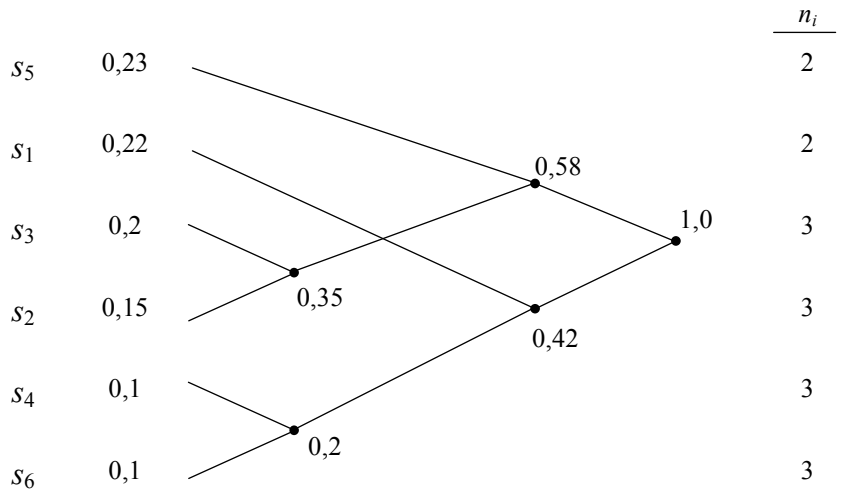
Teoria da Informação (2005-2006) Exame de 1ª chamada (28-6-06)

Resolução

1. Da tabela de contagens fornecida concluímos que as probabilidades de ocorrência dos diversos símbolos são as seguintes[1]: $P[s_1 \quad s_2 \quad \ldots \quad s_6] = [0,22 \quad 0,15 \quad 0,2 \quad 0,1 \quad 0,23 \quad 0,1]$.

a) Árvore da codificação de Huffman de variância mínima:



Logo, $\{n_1, n_2, \ldots, n_6\} = \{2,3,3,3,2,3\}$.

Comprimento médio das palavras de código: $\bar{N} = \sum_i n_i p_i = 2,55$ bits/símbolo.

Variância dos comprimentos: $V = E\left[\left(n - \bar{N}\right)^2\right] = E\left[n^2\right] - \bar{N}^2 = \sum_i n_i^2 p_i - \bar{N}^2 = 6,75 - 6,5025 = 0,2475$.

Nota: se a codificação não for de variância mínima obtêm-se $\bar{N} = 2,55$ (claro!) e $V = 0,6475$.

b) Entropia: $H(X) = H(p_1, p_2, \ldots, p_6) = -\sum_{i=1}^{6} p_i \log_2 p_i = 2,5076$.

Eficiência da codificação: $\frac{H(X)}{\bar{N}} = \frac{2,5076}{2,55} = 98,3\%$. (se tomarmos $H(X) = 2,51$ a eficiência vale 98,4%)

2. Código de repetição (5,1).

a) Só há duas palavras de código, 00000 e 11111. A matriz geradora é $\mathbf{G} = [11111]$ pelo que $\mathbf{P} = [1111]$ e

$$\mathbf{H} = \left[\frac{\mathbf{P}}{\mathbf{I}_{n-k}}\right] = \begin{bmatrix} 1111 \\ 1000 \\ 0100 \\ 0010 \\ 0001 \end{bmatrix}$$

b) Síndrome: $\mathbf{S} = \mathbf{ZH} = [11001]\mathbf{H} = [0110]$.

[1] Na verdade até nem precisaríamos de calcular as probabilidades… bastaria usar o número de ocorrências.

c) A matriz-padrão tem $2^{n-k} = 2^4 = 16$ "coset leaders" preenchidos com todos os padrões de 0, 1 e 2 erros assim distribuídos: 1 padrão de zero erros, $\binom{5}{1} = 5$ padrões de um erro e $\binom{5}{2} = 10$ padrões de dois erros. Verifica-se, portanto, que o limite de Hamming é atingido, $2^{n-k} = 1 + \binom{5}{1} + \binom{5}{2}$. Concluímos que o código de repetição (5,1) é perfeito e que $t = 2$ (esta última conclusão podia ter sido antecipada visto que $d_{\min} \geq 2t+1$ e $d_{min} = 5$).

Nota: além dos códigos de repetição de comprimento ímpar, como este, os únicos códigos binários perfeitos são os códigos de Hamming e o código de Golay (23,12).

d) Queremos calcular a probabilidade de erro não corrigido, $P_{enc} = 1 - \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i}$. Da matriz-padrão tiramos que $\alpha_0 = 1$, $\alpha_1 = 5$ e $\alpha_2 = 10$, $\alpha_3 = \alpha_4 = \alpha_5 = 0$. Logo,

$$P_{enc} = 1 - \sum_{i=0}^{2} \alpha_i p^i (1-p)^{n-i} = 1 - \left[(1-p)^5 + 5p(1-p)^4 + 10p^2(1-p)^3 \right] = 9{,}8.10^{-15} .$$

3. Segundo o enunciado os polinómios $45_8 = 100101$ ($p^5 + p^2 + 1$) e $51_8 = 101001$ ($p^5 + p^3 + 1$) são factores de $p^{31} + 1$ e também factores do polinómio gerador $g(p)$.

a) $g(p) = \left(p^5 + p^2 + 1 \right)\left(p^5 + p^3 + 1 \right) = p^{10} + p^8 + p^7 + p^5 + p^3 + p^2 + 1$. Concluímos que $n - k = 10$ e que $k = 21$.

b) Ao polinómio gerador corresponde uma palavra de código com peso 7. Logo, a distância mínima não pode ser 14. Poderemos concluir o mesmo determinando os majorantes da distância mínima dados pelos limites de Singleton e de Plotkin (aquele limite não é respeitado com $d_{\min} = 14$):

Singleton: $d_{\min} \leq n - k + 1 \quad \Rightarrow \quad d_{\min} \leq 11$

Plotkin: $d_{\min} \leq n \dfrac{2^{k-1}}{2^k - 1} \quad \Rightarrow \quad d_{\min} \leq 31 \times \dfrac{2^{20}}{2^{21} - 1} = 15{,}5$

c) Polinómio recebido: $p^{12} + p^{10} + p^7 + p^5 + p^4 + p^2$. A síndrome correspondente vale $S(p) = Z(p) \bmod g(p)$. Fazendo as contas chegamos a $S(p) = p^9$.

Com $g(p)$ alternativo ($g(p) = p^{10} + p^9 + p^8 + p^7 + p^6 + p^4 + p^3 + p^2 + 1$):

$$S(p) = Z(p) \bmod g(p) = p^9 + p^8 + p^7 + p^5 + p^2 + p + 1$$

d) Como $\mathbf{H} = \left[\dfrac{\mathbf{P}}{\mathbf{I}_{n-k}} \right]$ tem dimensões 31x10 a 21ª linha de **H** é a 21ª linha de **P**. Ora a linha de ordem i de **P** na forma polinomial é obtida calculando $p^{n-i} \bmod g(p)$, ou seja,

21ª linha de **H**: $p^{31-21} \bmod g(p) = p^{10} \bmod g(p) = p^8 + p^7 + p^5 + p^3 + p^2 + 1$ (0110101101).

(Com $g(p)$ alternativo: $p^{10} \bmod g(p) = p^9 + p^8 + p^7 + p^6 + p^4 + p^3 + p^2 + 1$ (1111011101))

4. Vamos interpretar o problema da seguinte forma:

- $X = \{A,B\}$ é o conjunto das apostas possíveis (feitas antes do jogo, claro); por exemplo, $P(X = A) = P_A = 0{,}6$ significa que a probabilidade de se apostar no país A é 60% (por outras palavras, o

apostador acredita que a equipa A tem 60% de probabilidades de ganhar). A entropia da "fonte" X vale $H(X) = \Omega(0,6) = 0,971$ bits/aposta.

- Ao fim do tempo regulamentar há três resultados possíveis, $Y = \{AV; AE; AD\}$ = {vitória de A; empate de A; derrota de A}, com $P(Y = AE) = \alpha$.
- $Z = \{A,B\}$ é o conjunto de resultados da final; por exemplo, $Z = A$ indica que o país A ganhou a Taça.

Note-se que $P(Y = AE \mid X = A) = P(Y = AE \mid X = B) = \alpha$ porque, de acordo com o enunciado, o prolongamento não depende nem de P_A nem de P_B, isto é, $P(Y = AE) = \alpha$, o que se pode escrever assim:

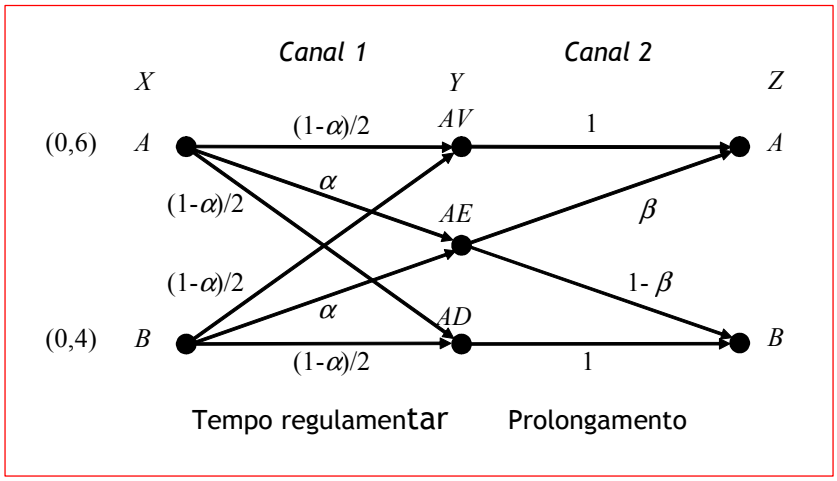
$$\begin{aligned} P(Y = AE) &= P_A P(Y = AE \mid X = A) + P_B P(Y = AE \mid X = B) = \\ &= P(Y = AE \mid X = A)\left(P_A + P_B\right) = \\ &= P(Y = AE \mid X = A) \end{aligned}$$

(veja-se o diagrama do canal abaixo para ajudar a clarificar a questão).

Como os resultados $Y = AV$ e $Y = AD$ ao fim do tempo regulamentar não dependem da aposta X então $P(Y = AV \mid X = A) = P(Y = AV \mid X = B)$ e $P(Y = AD \mid X = A) = P(Y = AD \mid X = B)$. Ora $\sum_i P(Y_i \mid X = A) = 1$ e $\sum_i P(Y_i \mid X = B) = 1$, logo

$$P(Y = AV \mid X = A) = P(Y = AV \mid X = B) = \frac{1-\alpha}{2}$$
$$P(Y = AD \mid X = A) = P(Y = AD \mid X = B) = \frac{1-\alpha}{2}$$

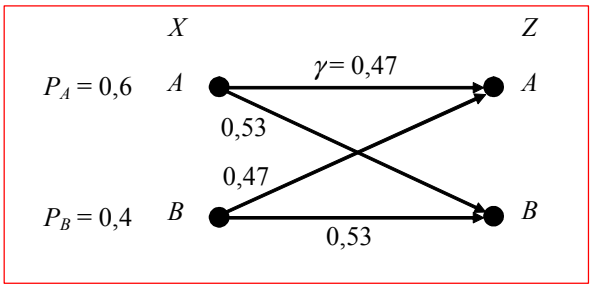
Podemos então considerar dois canais discretos, $X \to Y$ e $Y \to Z$, colocados em série como se mostra.



As matrizes de transição dos canais vão servir para obtermos o canal composto X → Z:

$$[Z \mid X] = \begin{bmatrix} \frac{1-\alpha}{2} & \alpha & \frac{1-\alpha}{2} \\ \frac{1-\alpha}{2} & \alpha & \frac{1-\alpha}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & 1-\beta \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha\beta + \frac{1-\alpha}{2} & \alpha(1-\beta) + \frac{1-\alpha}{2} \\ \alpha\beta + \frac{1-\alpha}{2} & \alpha(1-\beta) + \frac{1-\alpha}{2} \end{bmatrix}$$

Seja $\gamma = \alpha\beta + \frac{1-\alpha}{2}$. Substituindo valores obtemos um canal binário assimétrico $[Z \mid X] = \begin{bmatrix} 0,47 & 0,53 \\ 0,47 & 0,53 \end{bmatrix}$ com o diagrama seguinte:



a) Queremos determinar a entropia da situação do jogo ao fim do tempo regulamentar, ou seja, queremos $H(Y)$:

$$\begin{aligned} &P(Y = AV) = P_A \frac{1-\alpha}{2} + (1-P_A)\frac{1-\alpha}{2} = \frac{1-\alpha}{2} \\ &P(Y = AE) = \alpha \\ &P(Y = AD) = \frac{1-\alpha}{2} \end{aligned} \quad \Rightarrow \quad \begin{aligned} H(Y) &= H\left[\frac{1-\alpha}{2}, \alpha, \frac{1-\alpha}{2}\right] = \\ &= H\left(0{,}35; 0{,}3; 0{,}35\right) = 1{,}5813 \text{ bits/simbolo} \end{aligned}$$

b) Do gráfico do canal composto concluímos que a probabilidade de a equipa A ganhar a taça é

$$P(Z = A) = \gamma = 0{,}47$$

c) Deseja-se calcular a diminuição da dúvida sobre o resultado ao fim de 90 minutos por sabermos o resultado final, ou seja, queremos determinar $I(Y;Z) = H(Y) - H(Y \mid Z)$, que é o mesmo que $I(Y;Z) = H(Z) - H(Z \mid Y)$, uma fórmula mais conveniente. Da alínea anterior sabemos que $P(Z = A) = 0{,}47$ pelo que

$$H(Z) = \Omega(0{,}47) = 0{,}9974 .$$

Quanto a $H(Z|Y)$:

$$\begin{aligned} H(Z|Y) &= \sum_i P(Y_i) H(Z \mid Y_i) = \\ &= \frac{1-\alpha}{2}\underbrace{\Omega(1)}_{0} + \alpha\Omega(\beta) + \frac{1-\alpha}{2}\underbrace{\Omega(1)}_{0} = \alpha\Omega(\beta) = \\ &= 0{,}3\Omega(0{,}4) = 0{,}2913 \end{aligned}$$

Portanto, o "não-adepto" reduziu a dúvida em $I(Y;Z) = H(Z) - H(Z \mid Y) = 0{,}9974 - 0{,}2913 = 0{,}7061$ bits/símbolo.

d) Agora queremos determinar $I(X;Z) = H(Z) - H(Z \mid X) = H(X) - H(X \mid Z)$ e a equivocação $H(X \mid Z)$.

A entropia condicional $H(Z \mid X)$ vale

$$\begin{aligned} H(Z|X) &= \sum_i P(X_i) H(Z \mid X_i) = P_A\Omega(\gamma) + (1-P_A)\Omega(\gamma) = \\ &= \Omega(\gamma) = \Omega(0{,}47) \\ &= 0{,}9974 \end{aligned}$$

Concluímos que a informação mútua média é nula,

$$I(X;Z) = H(Z) - H(Z \mid X) = 0{,}9974 - 0{,}9974 = 0 \text{ bits/símbolo,}$$

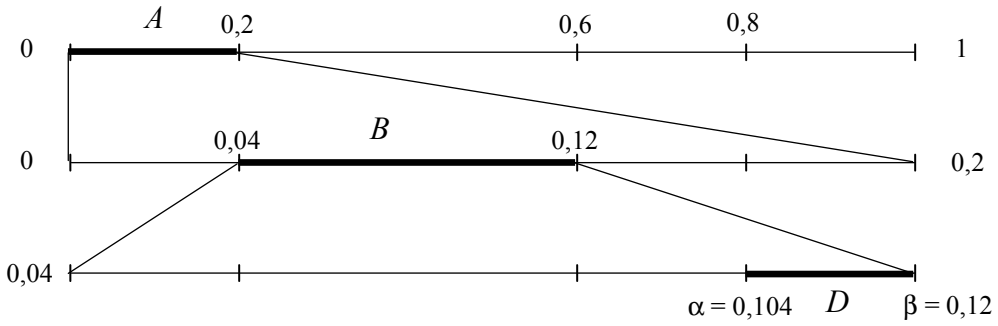
o que mostra que a aposta e o resultado final são independentes (como deviam).

A equivocação vale $H(X \mid Z) = H(X) - H(Z) + H(Z \mid X) = H(X) = 0{,}971$ bits/símbolo (igual à entropia da "fonte" X, claro, pois X e Z são independentes, como se viu).

e) A incerteza quanto ao vencedor é $H(Z) = \Omega(\gamma) = \Omega\left(\alpha\beta + \frac{1-\alpha}{2}\right)$ e não depende em nada da probabilidade P_A. Portanto, qualquer que seja o prognóstico, a incerteza quanto ao vencedor é a mesma (como devia ser!).

5. Codificação aritmética de ABD, com $P(A) = 0{,}2$, $P(B) = 0{,}4$, $P(C) = 0{,}2$ e $P(D) = 0{,}2$.

a) Obtenção do intervalo final de codificação:



Daqui se conclui que o intervalo final é [α; β[= [0,104; 0,12[e a sua largura é $L = \beta - \alpha = 0{,}016$.

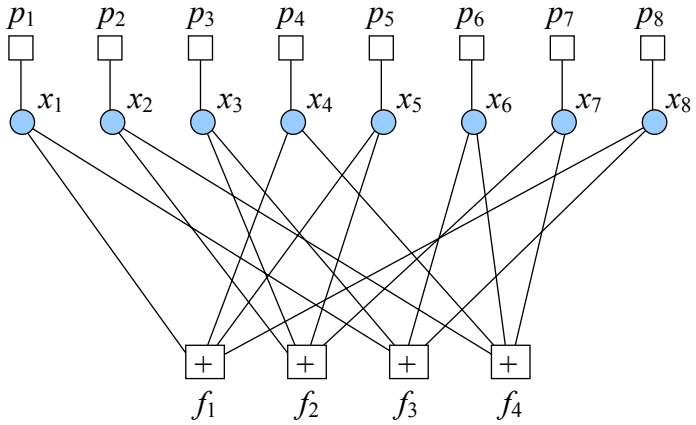
b) Qualquer valor real do intervalo [0,104; 0,12[serve (pode é não ser o mais conveniente se quisermos obter a palavra binária mais curta…). Assim, um valor possível é 0,11.

c) Queremos obter a fracção diádica $x/2^t$ com o menor denominador:

- $t = \lceil -\log_2 L \rceil = 6$
- $\alpha \le \frac{x}{2^t} < \beta \quad \Rightarrow \quad x < 2^t \beta = 64 \times 0{,}12 = 7{,}68$ e $x \ge 2^t \alpha = 64 \times 0{,}104 = 6{,}66$, isto é, $x = 7$.

A fracção diádica procurada é $\frac{7}{2^6}$ (igual a 0,1094), a que corresponde a palavra binária 000111 (converte-se o numerador 7 numa palavra binária de seis bits).

6. a) Gráfico de Tanner (incluindo o canal de comunicação):



b) E_b/N_0 (dB) = 1 dB corresponde a $E_b/N_0 = 1{,}259$. Como $R_c = 1/2$ então a variância do ruído vale

$$\sigma^2 = \frac{1}{2R_c\, E_b/N_0} = \frac{1}{1{,}259} = 0{,}79$$

As probabilidades a priori (assinaladas no gráfico de Tanner acima) são dadas por

$$p_i = p(y_i \,|\, x_i = +1) = \frac{1}{1+\exp\left(-2y_i/\sigma^2\right)}$$

Substituindo nesta expressão os valores da sequência recebida

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_8 \end{bmatrix} = \begin{bmatrix} \text{-1,38} & \text{-2,48} & 1{,}11 & 1{,}26 & \text{-2,02} & 2{,}06 & 2{,}06 & \text{-1,03} \end{bmatrix}.$$

encontramos as probabilidades que nos interessam nesta alínea:

$$p_2 = 0{,}002 \qquad p_4 = 0{,}960 \qquad p_5 = 0{,}006 \qquad p_6 = 0{,}994$$

Logo,

$$\begin{aligned} q_{52} &= p_5 = 0{,}006 \\ r_{47} &= q_{24}(1-q_{44})(1-q_{64}) + q_{44}(1-q_{24})(1-q_{64}) + q_{64}(1-q_{24})(1-q_{44}) + q_{24}q_{44}q_{64} = \\ &= 4{,}8.10^{-7} + 0{,}0057 + 0{,}0397 + 0{,}0019 = \\ &= 0{,}0473 \end{aligned}$$

c) Algoritmo *max-product*.

- a mensagem q_{52} é a mesma do algoritmo não simplificado: $q_{52} = 0{,}006$.
- a mensagem do nó de paridade é igual à maior parcela da soma na expressão de r_{47}: $r_{47} \approx 0{,}0397$.

(Isto não foi pedido) Vejamos como utilizar o algoritmo *min-sum*, se o desejássemos. Nesse caso usam-se LLR e só as mensagens dos nós de paridade são calculadas de maneira simplificada. Assim, sendo $L(p_i) = 4\frac{R_c E_b}{N_0} y_i = 2{,}518 y_i$ temos

$$\begin{aligned} L(q_{52}) &= L(p_5) = 2{,}518 \times (-2{,}02) = -5{,}09 & \qquad L(q_{44}) &= L(p_4) = 3{,}173 \\ L(q_{24}) &= L(p_2) = \text{-6,244} & \qquad L(q_{64}) &= L(p_6) = 5{,}187 \end{aligned}$$

Quanto a $L(r_{47})$, no algoritmo *min-sum* é igual a

$$L(r_{47}) = (-1)^{d_4} \left(\prod_{i' \in \{2,4,6\}} \operatorname{sgn}\left[L(q_{i'4})\right] \right) \min_{i' \in \{2,4,6\}} \left(\left| L(q_{i'4}) \right| \right).$$

Neste caso concreto o nó 4 tem grau $d_4 = 4$ e, portanto, $L(r_{47}) = (-1)^4 (-1) \min\left(6{,}244; 3{,}173; 5{,}187\right) = -3{,}173$.

Nota: sem aproximação a mensagem vale $L(r_{47}) = 2\tanh^{-1}\left[-\prod_{i' \in \{2,4,6\}} \tanh\left(-L(q_{i'4})/2\right) \right] = -3{,}008$.

d) Queremos $L(q_{13}) = L(r_{11}) + L(p_1)$.

$$\begin{aligned} L(r_{11}) &= \ln\frac{r_{11}}{1-r_{11}} = \ln\frac{0{,}289}{0{,}711} = -0{,}900 \\ L(p_1) &= 4\frac{R_c E_b}{N_0} y_1 = 2{,}518 y_1 = 2{,}518 \times (-1{,}38) = \text{-3,475}. \end{aligned}$$

Portanto, $L(q_{13}) = L(r_{11}) + L(p_1) = -0{,}900 - 3{,}475 = \text{-4,375}$.